

# Understanding consumer demand in customised pricing environments.

Shreena Patel, M.A.(Hons), M.Res.



Submitted for the degree of Doctor of Philosophy  
at Lancaster University.

March 2016

# Abstract

Estimation of price sensitivity from real world data is typically complicated by a dependence between price and demand, or endogeneity of price, since prices are regularly varied in anticipation of changes to demand. This problem is particularly severe in customised pricing environments, where sellers have some freedom to quote different prices for different orders, based on information about each customer and their order. When endogeneity is left untreated and price is modelled as an independent predictor, this leads to underestimation of price sensitivity and sub-optimal pricing strategies.

Endogeneity bias is corrected by the inclusion of instrumental variables in the model; these are variables which are correlated with price and independent of demand, and allow us to separate the direct and indirect effects of price on demand. Whilst instrumental variable estimation has been well documented for retail pricing problems, their use in customised pricing is relatively under-represented in marketing literature.

Here we present a probit model of purchasing behaviour for these environments, whereby the price offered to a customer and their corresponding price threshold are represented by a bivariate Gaussian random variable. Recorded or known sources of dependence between the two are introduced via covariate effects on the mean and

unrecorded sources of dependence are captured by a residual correlation parameter. The parameters are identified by the inclusion of instrumental variables in the model. Using a two-stage estimation procedure, the model is fitted to telesales data for heating oil, and price sensitivity estimates are compared to those of a naive model, which does not correct for endogeneity bias. Bayesian estimation of the model is then performed via MCMC, and the resulting sample of parameters are used to examine the impacts of various price changes on profit.

# Acknowledgements

Firstly I would like to thank my supervisor, Dr. Chris Sherlock, for his guidance and support throughout the past years. His encouragement and understanding were invaluable and helped me to keep going during the tougher parts of my research. I am also grateful to Prof. Paul Fearnhead for his excellent insight into the project.

My PhD was supported by EPSRC and Kalibrate Technologies plc, as part of the Statistics and Operational Research (STOR-i) doctoral training centre. I could not have asked for a more supportive or stimulating research environment and I am grateful to STOR-i for the many learning opportunities it offered. Special thanks goes to Chris Nemeth, Jamie Fairbrother, Tim Park and Mark Bell for their advice and friendship, and for all the fun we had over the past years.

Finally, I would like to express my gratitude to my family for their continuous encouragement, patience and support. I am proud to have two of the strongest and most selfless people I know as a mother and sister. I am forever grateful to them for their belief in me.

# Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

Shreena Patel

# Contents

<b>Abstract</b>	<b>I</b>
<b>Acknowledgements</b>	<b>III</b>
<b>Declaration</b>	<b>IV</b>
<b>Contents</b>	<b>V</b>
<b>List of Figures</b>	<b>VI</b>
<b>List of Tables</b>	<b>VII</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Customised pricing . . . . .	3
1.2 Endogeneity and instrumental variables . . . . .	6
1.3 Heating oil pricing . . . . .	8
1.4 Thesis outline . . . . .	9
<b>2 Background</b>	<b>11</b>
2.1 Customised pricing . . . . .	11

2.1.1	Price optimisation . . . . .	13
2.1.2	Estimation of willingness to pay . . . . .	16
2.1.3	Updating and monitoring . . . . .	20
2.2	Endogeneity . . . . .	22
2.2.1	Instrumental variable estimation . . . . .	23
2.3	Review of literature on the use of IV estimation in pricing . . . . .	29
2.4	Bayesian estimation . . . . .	34
<b>3</b>	<b>Modelling willingness to pay</b>	<b>40</b>
3.1	Formulation . . . . .	40
3.2	Price elasticity . . . . .	43
3.3	Parameter estimation . . . . .	47
3.3.1	Identifiability . . . . .	48
3.3.2	Two-stage estimation . . . . .	49
3.3.3	Willingness to pay without endogeneity . . . . .	52
3.4	Estimation of standard errors . . . . .	53
3.4.1	Standard errors of parameter estimates . . . . .	53
3.4.2	Standard errors of predictions . . . . .	56
3.5	Simulated example . . . . .	57
3.6	Modelling considerations . . . . .	64
3.6.1	Treatment for non-normality . . . . .	64
3.6.2	Measuring goodness of fit . . . . .	66
3.6.3	Non-linear covariate effects . . . . .	67

3.6.4	Random effects . . . . .	69
<b>4</b>	<b>Heating Oil Case Study</b>	<b>74</b>
4.1	The market for heating oil . . . . .	74
4.1.1	Description of dataset and data cleaning . . . . .	76
4.1.2	Selection of price and cost variables . . . . .	79
4.1.3	Exploratory data analysis . . . . .	80
4.1.4	Empirical willingness to pay . . . . .	90
4.2	Modelling procedure . . . . .	92
4.2.1	Heteroscedasticity . . . . .	92
4.2.2	Non-normality . . . . .	96
4.2.3	Model selection . . . . .	98
4.3	Results . . . . .	101
4.4	Discussion . . . . .	103
4.4.1	Covariate effects . . . . .	103
4.4.2	Random effects . . . . .	106
4.4.3	Instrumental variables . . . . .	109
4.4.4	Estimates of price sensitivity . . . . .	112
4.5	Summary . . . . .	116
<b>5</b>	<b>Profit simulation</b>	<b>118</b>
5.1	Introduction . . . . .	118
5.1.1	Predicting willingness to pay . . . . .	121
5.1.2	Measuring profit improvement . . . . .	124



5.2	Bayesian estimation . . . . .	126
5.2.1	Recap of model . . . . .	126
5.2.2	Prior and posterior distributions . . . . .	129
5.2.3	MCMC scheme . . . . .	131
5.2.4	Chain diagnostics . . . . .	134
5.2.5	Parameter estimates . . . . .	137
5.3	Counterfactual analysis . . . . .	141
5.3.1	Expected profit as a model check . . . . .	141
5.3.2	Predicting profit improvement . . . . .	143
5.3.3	Alternative pricing policies . . . . .	147
5.3.4	Results . . . . .	150
5.3.5	Discussion of results . . . . .	153
5.4	Summary . . . . .	157
<b>6</b>	<b>Conclusions</b>	<b>159</b>
6.1	Final remarks and contributions . . . . .	159
6.2	Directions for future work . . . . .	161
<b>A</b>	<b>Parameter estimates</b>	<b>162</b>
	<b>Bibliography</b>	<b>168</b>

# List of Figures

3.2.1 Direct (a) and indirect (b) effects of price on willingness to pay. The shaded areas represent the probability of sale. (a) Given that the customer was offered a price of $q$ , an increase in the hypothetical price from $\tilde{q}_1$ to $\tilde{q}_2$ causes the probability of sale to decrease from the lighter region to the darker region. (b) If the agent were to offer a quote of $q_2$ rather than $q_1$ this would imply a different referral price mean and therefore a shift in the referral price distribution. Hence for a given hypothetical price, $\tilde{q}$ , the probability of sale is higher under $q_2$ than $q_1$ .	45
3.5.1 Simulated example of quote and referral price as a function of unobserved covariate $x_c$ , resulting in an upward-sloping WTP curve (right). A (0,1) line is included in the centre plot for reference (-----).	59
3.5.2 The relationship between willingness to pay and the estimated quote price residuals from each model.	62
3.5.3 Predictions of willingness to pay for a new sample, alongside observed acceptance rates for the IV and full model.	63

4.1.1 Daily average retail price, wholesale price and spot price of kerosene over the study period. . . . .	75
4.1.2 Different representations of the relationship between price and cost, based on retail price, wholesale price and spot price, with a line of best fit (—). . . . .	80
4.1.3 The median (—) and upper and lower quartiles (----) of margins offered on each day of business. . . . .	81
4.1.4 Changes in the average daily wholesale price and corresponding changes in average retail price, with a line of best fit (—) and (0, 1) line for reference (----). . . . .	81
4.1.5 Average features for each depot. Points are colour coded by depot and in the left and right plots, point size is proportional to the number of quotes requested for each depot. In the centre plot, point size is proportional to average margin for each depot. . . . .	82
4.1.6 Total volume of fuel sold per day at one of the largest (—) and smallest (—) depots. . . . .	83
4.1.7 Approximate depot locations, where point size is proportional to the no. of customers for each depot. . . . .	84
4.1.8 Average margin for each postcode on 27 <sup>th</sup> Feb 2006, the busiest day of the study period. . . . .	84
4.1.9 Relationship between <code>dist</code> and <code>margin</code> (left) and <code>dist</code> and <code>acc</code> (right). Observations have been binned according to quantiles of <code>dist</code> . . . . .	84

4.1.10	Average margin and acceptance rates (with 95% confidence intervals) across orders of different sizes. Quotes were divided into bins according to volume. . . . .	85
4.1.11	Relationship between volumes requested by customers and tempera- ture. Quotes were divided into bins according to temperature. . . . .	86
4.1.12	Average margin and acceptance rates (with 95% confidence intervals), according to the number of quotes the customer has previously requested. . . . .	87
4.1.13	Average margin and acceptance rates (with 95% confidence intervals) across orders at different stages of the customer's purchase history. Categories have been grouped for easier interpretation: one or no pre- vious quotes, two previous quotes, three previous quotes and four or more previous quotes. . . . .	88
4.1.14	Relationship between <code>quoteInt</code> and <code>margin</code> (left) and <code>quoteInt</code> and <code>acc</code> (right). Observations have been binned according to the number of days since the customer last requested a quote. . . . .	89
4.1.15	Proportion of quotes accepted for different intervals of margin, or $q$ . . . . .	91
4.2.1	Kernel density plots of residuals from initial quote price model, cate- gorised by depot. . . . .	94
4.2.2	Hierarchical clustering of residual variances for each depot, illustrated as a dendrogram (left) and a corresponding elbow plot (right). . . . .	95
4.2.3	QQ plots of residuals from the quote price model for each subset, both (a) before and (b) after a Box-Cox transformation. The 1 <sup>st</sup> , 5 <sup>th</sup> , 96 <sup>th</sup> and 99 <sup>th</sup> sample percentiles are shown by dashed lines. . . . .	96

4.2.4 Kernel density plots of residuals from the final quote price model, where the sample was divided into three subsets based on depot. Residuals are grouped by depot and a colour key is provided in Figure 4.2.1. . . .	97
4.2.5 Observed vs. expected acceptance rates, for the WTP model with price as margin (left) and price as the Box-Cox transformed margin (right). .	98
4.2.6 BLUPs of random effects in the quote price model versus effects in the referral price model. . . . .	100
4.3.1 MLEs and 95% confidence intervals for the quote price parameters, $\beta_q$ . The intercepts for Subsets 1-3 are 2.87, 2.72 and 2.20 respectively. . . .	101
4.3.2 MLEs and 95% confidence intervals for the referral price parameters, $\beta_r$ . The intercepts for Subsets 1-3 are -2.99,-3.17 and 0.31 respectively.	102
4.4.1 Average customer price effect against agent price effect. . . . .	108
4.4.2 Average customer price effect for each postcode. . . . .	108
4.4.3 Relationship between the customer price effect, <code>custEff</code> , and <code>acc</code> . Observations have been binned according to <code>custEff</code> . . . . .	109
4.4.4 Relationship between cost and margin (left) and cost and observed acceptance rates (right) . . . . .	110
4.4.5 The instrumental variables, <code>month</code> and <code>cost</code> , were each excluded from the quote price model for Subset 1 and are plotted here against the residuals. Left: average residual for each month, with 95% confidence intervals (—). Right: average residual for different intervals of cost, with 95% confidence intervals (-----). . . . .	111

4.4.6 The difference between average and predicted acceptance rates, based on the referral price model, alongside 95% confidence intervals. Observations are grouped by month (left) and for different intervals of cost (right). . . . .	112
4.4.7 Empirical willingness to pay curves for each subset (top) and the relationship between quote price residuals and willingness to pay. . . . .	114
5.2.1 Point estimates and standard errors for fixed effects parameters and random effects from the quote price model for Subset 3. . . . .	127
5.2.2 Example traceplots for the first of the three runs of the Markov chain. The top row shows three of the referral price parameters, $\phi_r$ , and the bottom row shows three of the customer random effects, $\mathbf{b}_c$ . . . . .	135
5.2.3 Combined Gelman-Rubin statistics for all 53 referral price parameters, $\phi_r$ (left) and all 9,834 customer random effects, $\mathbf{b}_c$ . The ideal statistic of 1 (—) and the threshold of 1.2 (—) are also shown. . . . .	135
5.2.4 Example autocorrelation plots for the first of the three runs of the Markov chain. The top row shows three of the referral price parameters, $\phi_r$ , and the bottom row shows three of the customer random effects, $\mathbf{b}_c$ . . . . .	136
5.2.5 Example kernel density plots for the posterior samples of parameters. The top row shows three of the referral price parameters, $\phi_r$ , and the bottom row shows three of the customer random effects, $\mathbf{b}_c$ . Corresponding MLEs and BLUPs are shown by a vertical line (—) . . . . .	138

5.2.6 Best linear unbiased predictors of the customer random effects for price, $\hat{\mathbf{b}}_c$ , plotted against corresponding posterior means, with a (0,1) line for reference (-----). . . . .	139
5.2.7 Posterior means and 95% HPD intervals for the raw referral price pa- rameters, $\phi_r$ . ML estimates are shown for comparison. . . . .	140
5.3.1 Kernel density plots for posterior samples of profit improvement after a 1% increase in retail prices (left) and a 1% decrease in prices (right). . . . .	151
5.3.2 Kernel density plots for posterior samples of profit improvement af- ter applying differentiated (left) and uniform (right) price changes to Depots 240 and 260. . . . .	152

# List of Tables

3.5.1 Estimates of mean parameters for simulated example . . . . .	60
3.5.2 Estimates of variance parameters for simulated example . . . . .	61
4.1.1 Variable definitions . . . . .	77
4.2.1 Data subsets . . . . .	95
4.4.1 Model contributions ( $\Delta$ BIC) of each covariate. . . . .	103
4.4.2 Decomposing quote price variability. . . . .	106
4.4.3 Williness to pay parameters . . . . .	113
4.4.4 Willingness to pay parameters under independence . . . . .	115
5.2.1 Numbers of random effects and parameters in WTP model for Subset 3	128
5.2.2 Effective sample sizes . . . . .	137
5.3.1 Expected profit under original prices . . . . .	143
5.3.2 Details of pricing policies . . . . .	149
5.3.3 Posterior estimates of profit improvement (£) after universal price changes	150
5.3.4 Variance decomposition for profit improvement (£) after universal price changes . . . . .	151



5.3.5 Posterior estimates of profit improvement (£) after price changes for each depot . . . . .	152
5.3.6 Posterior estimates of profit improvement (£) after differentiated and uniform price changes for Depots 240 and 260 . . . . .	152
5.3.7 Posterior estimates of profit improvement (£) after price changes based on <code>quoteInt</code> . . . . .	153
5.3.8 Posterior estimates of profit improvement (£) after price changes based on <code>quoteHist</code> . . . . .	153
5.3.9 Posterior estimates of profit improvement (£) after price changes based on $\mathbf{b}_c$ . . . . .	153
A.0.1 Quote price parameter estimates . . . . .	162
A.0.2 Referral price parameter estimates . . . . .	165

# Chapter 1

## Introduction

Endogeneity occurs when there is an additional relationship between a covariate and a response, above and beyond the direct relationship that is of interest. For example, suppose that another unobserved variable affects both the observed covariate and the response. The apparent relationship between the observed covariate and the response will also account for this hidden relationship, and any naive estimator of the direct relationship of interest is likely to be biased. This frequently arises in observational studies and has recently received attention in customised pricing, where prices are varied from transaction to transaction based on information about each customer and their order. As a result, the prices offered to customers are not independent of their willingness to pay and standard methods of estimating price sensitivity will be biased and inconsistent.

Instrumental variable techniques are widely used in econometric models to avoid the endogeneity bias of including an independent variable that is correlated with the error term. This thesis is centred around the use of instrumental variables in

customised pricing. We develop a probit model for willingness to pay in which the seller's pricing process and the customer's purchase decision are modelled jointly by a bivariate Gaussian distribution. Observed sources of dependence are represented by covariate effects on the mean and any remaining dependence is captured by residual correlation. The model can be estimated in two stages; firstly a linear model is fitted to the prices offered by the seller. The residuals from this model are then entered into a probit regression of the customer responses against price and other available covariates. Provided that an instrumental variable is available - that is, a covariate which affects price but not willingness to pay - the model will be identifiable.

By conditioning on observed sources of variation in price, we can separate the direct effect of price on willingness to pay from the indirect effect caused by residual correlation. This provides an estimate of true price sensitivity, which can then be used to predict willingness to pay under new prices. As a motivating example, we examine sales data for heating oil, in which agents have been given some degree of pricing discretion. The acceptance rate for price offers is observed to increase with price, as a result of dependence between price and willingness to pay. The bivariate model is developed for this dataset in order to correct for endogeneity bias, and allows us to better identify the causal relationship between price and willingness to pay. Based on these results, we then examine the impact of price changes to the seller's profit. Bayesian estimation of the model is performed to obtain posterior samples of the model parameters, which are then used to predict profit under a new set of prices. In this way, we can identify opportunities for the seller to improve profit.

## 1.1 Customised pricing

The past fifty years has seen demand management emerge as both a popular business practice and an active area of research within the economics and marketing literature. Now an integral part of the hotel, retail and car rental industries, a large proportion of the work in this area takes place in a list price setting. Here the seller commits to offering a price for a fixed period of time, before learning the identity of any potential customers. More recent efforts at price optimisation focus on customised selling environments, in which prices are offered on a transaction-by-transaction basis. The flexibility to adjust prices in real-time, according to market conditions and individual customer behaviour, provides considerable potential for demand management to improve pricing decisions in this setting. United Parcel Service, for instance, reported an annual profit improvement of over \$100 million by optimising their customised pricing system (Boyd et al., 2005).

Whilst price discrimination is as old as trade itself, the use of statistical models for price discrimination has only become feasible with the advent of modern technologies. Customer resource management (CRM) systems allow sellers to capture and store information about millions of transactions and access this information instantly, at the individual level, while responding to a customer's enquiry. Rapid growth in computational capabilities has made it possible to analyse this data, execute complex optimisation algorithms and then implement and manage the resulting pricing decisions. However despite its growing prevalence in the economy, customised pricing has been relatively less studied as a pricing mechanism.

List pricing concerns *demand*, or the quantity of a product purchased by customers, whereas customised pricing concerns *willingness to pay*, the probability of a customer accepting a particular price offer. In the latter case, sellers are able to observe instances where a customer receives a price offer and chooses not to purchase, termed *lost sales*. Consumer lending, insurance and business-to-business trade are particularly well-suited to price customisation as sellers often gain detailed information about the customer before quoting a price. In addition to customer characteristics, such as location and purchase history, the seller is aware of the product(s) requested and the channel through which the buyer approached the seller to request a quote. These markets are typically characterised by *monopolistic competition*, whereby a seller maintains some degree of control over the price they set, since their products are not perfect substitutes with another competitor's. Hence the seller has the freedom to quote different prices to different customers based on the information available to him.

Key to customised pricing is customer heterogeneity, or variability in the amount customers are willing to pay for a product. Services and highly customisable goods tend to fall under this category, as the value of the product to the customer is intangible to a large extent. In making a purchase decision, customers are likely to consider some combination of 'objective' factors, such as price and 'subjective' factors, such as the brand image of the seller and the customer's experience in previous dealings with them (Phillips, 2005). For this reason, the seller quoting the lowest price is not guaranteed to win the customer's business, and a seller may command a positive or negative price premium over his competitors. In other words, a customer may be

willing to pay more or less to purchase a product from a particular seller, compared to other sellers operating in the market.

With perfect information about the price premium a customer is willing to pay for his product, a seller could use one-to-one pricing to maximise the profitability of an individual transaction. However in practice, the seller's pricing strategy is subject to both *preference uncertainty* and *competitive uncertainty* (Phillips, 2010). The former refers to the exact criteria used by the customer to compare quotes and the latter arises because the seller is typically unaware of the prices quoted by his competitors. Price optimisation systems use statistical inference to recommend an optimal price under this imperfect information.

In the absence of an analytical pricing scheme, the majority of sellers rely on a team of sales agents to discriminate between different customers. This allows the seller to respond to a large volume of enquiries, whilst also taking advantage of any incremental profit gains that arise. Often sales agents have restricted pricing flexibility, whereby a centralised pricing authority provides a price recommendation, which agents are not strictly required to adhere to. Hence agents can use their discretion to vary prices across customers and also change the price of a product for the same customer over time.

It can be argued that sales agents are in the best position to assess market conditions and adjust prices accordingly, however agents are not necessarily risk-neutral and their pricing decisions are often biased and sub-optimal (Elmaghraby et al., 2012). In particular, agents are typically compensated by sales volume and therefore inclined to offer prices below their profit-maximising levels. Evidence has shown that aligning

sales agents' incentives with those of senior management, for example by monitoring compliance with pricing recommendations, tends to improve business performance (Garrow et al., 2006). Analytical approaches to customised pricing aim to present an unbiased price recommendation for a particular transaction.

## 1.2 Endogeneity and instrumental variables

Price optimisation systems typically maximise the expected profit of a transaction, which requires an estimate of the customer's willingness to pay; that is, the probability of the customer accepting a particular price given the information available to the seller at the time. Willingness to pay is predicted from historical transaction data. In the standard case, the customer responses (accept or reject) are regressed against price and any other relevant variables recorded for each transaction. However in a customised selling environment, price cannot be treated as an independent regressor.

The seller's flexibility to vary prices over time and across customers often leads to a correlation between price and demand in transaction data; for example, if a seller anticipates low demand over a particular period, they may lower prices in order to attract customers. This correlation is caused by their joint dependence on a common set of factors, such as supply shocks or customer location. To the extent that these factors are observable, we can account for the dependence between price and demand and obtain unbiased estimates of price sensitivity.

*Endogeneity* refers to the case where dependence between price and demand is caused by unobservable factors, such as negotiation between the seller and customer.

This often occurs in a customised price setting, as agents are given pricing discretion. The price offered to a customer tends to be correlated with their probability of acceptance, since the agent will increase or decrease the price according to her interaction with the customer. As a result, sales data from customised selling environments tends to display less variability in customer responses than we would observe under independence between price and demand.

Untreated, endogeneity gives rise to underestimates of price sensitivity and sub-optimal pricing recommendations, as it obscures the causal relationship between price and demand. Within the econometrics literature, a popular solution for endogeneity is to include *instrumental variables* in the model; these are extra variables which are correlated with price and independent of demand. The role of the instrument is to inject variation into prices which is independent of demand and in doing so, provide unbiased estimates of price sensitivity.

Endogeneity is an inherent feature of customised pricing, as sellers have comprehensive information about the customer before offering them a price, and data is collected at the individual customer level rather than aggregate level. In list pricing, the primary source of endogeneity is the ability of the seller to anticipate seasonal changes in total demand. Whereas endogeneity in customised pricing is caused by the ability of the seller (or agent) to infer latent customer variables. For this reason, endogeneity tends to be a more severe problem in customised selling environments and correction is less straightforward. The measurement and treatment of endogeneity in list pricing has been well documented (Petrin and Train, 2010; Berry et al., 1995; Hormby et al., 2010; Villas-Boas and Winer, 1999; Yang et al., 2003; Park and



Gupta, 2009), whilst only a handful of papers address endogeneity in the context of customised pricing (Phillips et al., 2012; Zhang et al., 2014; Elmaghraby et al., 2012).

### 1.3 Heating oil pricing

The majority of off-grid households (those not served by natural gas pipelines) rely on the delivery of heating oil (or kerosene) to heat their homes. When a customer calls a supplier to request a quote they are offered a price and delivery date for the quantity of fuel they wish to buy, which they can then accept or reject. Currently, the agent uses a customer relationship management system to determine an appropriate price to quote. This system gives the agent information about the customer's sales history and a reference price, which the agent adjusts upwards or downwards depending on their assessment of the customer's price sensitivity, short term fluctuations in demand and current capacity limitations.

Although heating oil is a commodity, the market exhibits many of the properties of a customised pricing environment. Customer heterogeneity, in the form of location and purchase history, as well as volatility in the wholesale price means that prices offered to different customers over different periods of time are rarely comparable. Pricing discretion on the part of sales agents often leads to dependence between price and demand, and the resulting pricing process is complex and inconsistent.

Kalibrate Technologies provide a consultancy service to suppliers in this business and wish to develop a system whereby each agent will be given a guide price to offer the customer. This price will have been chosen to optimise profits subject to constraints

on the delivery of fuel, using predictions of the amount a customer is willing to pay. The aim of this research is to provide estimates of willingness to pay which can be used as a basis for a more formal pricing policy. Since price is an endogenous regressor within this context, endogeneity correction is required to obtain accurate estimates of customers' price sensitivity.

## 1.4 Thesis outline

The main material of this thesis is divided into four chapters: an introduction and literature review of the area (Chapter 2); the model developed for the pricing problem (Chapter 3); a case study into heating oil pricing (Chapter 4); and a simulation study based on the model results (Chapter 5). A brief outline for each chapter is provided below:

### **Chapter 2: Background**

This chapter provides some background information on instrumental variables and the customised pricing process. In particular, current literature on the use of instrumental variable estimation in pricing is reviewed. An introduction to Bayesian estimation is also provided, which forms the basis of the simulation study in Chapter 5.

**Chapter 3: Willingness to pay model**

A bivariate normal model representing the seller's pricing process and the customers' buying process is presented and used to derive a probit function for willingness to pay. The identifiability of the model is discussed, as well as the estimation of standard errors. A simple example is presented with simulated data, and extensions of the model are put forward.

**Chapter 4: Heating oil case study**

This chapter describes the estimation of the willingness to pay model for the heating oil sales data. Several covariates are introduced and model fitting choices regarding transformations and segmentations of the data are discussed. Random effects are introduced to account for customer heterogeneity, and the independence and relevance of the selected instrumental variables is validated. The model is evaluated by comparison to a standard model, where price is treated as an independent regressor.

**Chapter 5: Profit simulation**

In this chapter, the impact of price changes on profit are evaluated by Bayesian estimation of willingness to pay. The model developed for the heating oil data is fitted by Adaptive MCMC, and the resulting posterior sample of parameters is used to predict profit. A counterfactual analysis is conducted whereby historical prices are modified, both over the entire sample and for targeted customer segments, in order to evaluate the impact on profit.

The thesis concludes with a summary of the contributions made by this study and some suggestions for future work.

# Chapter 2

## Background

Below we present an introduction to the customised pricing process and some background information on instrumental variables; this is followed by an overview of the literature in the area and an introduction to Bayesian estimation. Customised pricing can be broken down into three main stages: estimation of willingness to pay, price optimisation, and updating and monitoring. Instrumental variables are involved in the estimation of willingness to pay when endogeneity correction is required. We present the typical situations in which endogeneity arises and define instrumental variable estimators for linear models.

### 2.1 Customised pricing

Let  $q$  and  $c$  denote the unit price and unit cost of a product and let  $V$  be the volume of the order. The function  $h(q, c, V)$  returns the profit from the sale,

$$h(q, c, V) = V \cdot (q - c), \quad (2.1.1)$$

and is assumed to take this form throughout the thesis. The *willingness to pay* function,  $\pi(q; \mathbf{x})$ , is the probability of the customer accepting a price of  $q$ . Since there is heterogeneity in demand, both over time and across customers, willingness to pay is also a function of the non-price attributes of the sale, denoted by  $\mathbf{x}$ .

In the most general case, the optimal price of a sale can be expressed as follows:

$$q^* = \operatorname{argmax}_q h(q, c, V) \times \pi(q; \mathbf{x}). \quad (2.1.2)$$

The optimal price,  $q^*$ , maximises the objective function, which is typically the expected profit from the sale. Since the above optimisation applies no constraints to  $q$ , it is known as the *unconstrained* price optimisation problem. In certain contexts, constraints on prices may be necessary. For example, the seller may wish to impose upper and lower bounds on price,  $q^- \leq q \leq q^+$ , in order to maintain some degree of price stability. Section 2.1.1 considers price constraints arising from the delivery of heating oil.

Estimation of the willingness to pay function is the first and arguably most crucial aspect of a customised pricing system, as its purpose is to represent any differences in customers' price sensitivity that the seller can use to vary prices. Customer responses in historical sales data are used to predict how likely the customer is to accept a quote. With this information, we can find the best price to offer the customer given the

information available to the seller. In the absence of any constraints, this calculation can be trivial; however with multiple products and several time-related constraints, the problem is significantly more complex. Since pricing decisions need to be made in real time, optimisation procedures must be executable within minutes.

A third aspect of customised pricing is the monitoring and updating of the willingness to pay function, since these are estimated under the assumption that past patterns of demand provide an accurate reflection of current and future demand. Any changes to competitors' pricing behaviour, customer preferences or macroeconomic conditions will require pricing policies to be reviewed. We discuss these stages in further detail below.

### 2.1.1 Price optimisation

The profit-maximising price optimally balances the effects of raising the price to increase profit and lowering the price, in order to increase the probability of the quote being accepted. For most business transactions, the profit from the sale is simply price minus cost,  $(q - c)$ , or alternatively  $V \cdot (q - c)$ , if price and cost are expressed in unit terms and  $V$  represents the volume of the order. In some markets the seller may face uncertainty regarding the cost of the order; for example in consumer credit, where cost represents the risk of default, in which case the profit function expresses the expected profit from the sale.

A universal measure of price sensitivity is *price elasticity*,  $\nu(q; \mathbf{x})$ , which measures the responsiveness of demand to a change in price. Strictly speaking, this is a local estimate of price sensitivity, found by taking the limit of *arc elasticity*. Arc elasticity,

$\nu_A(q_1, q_2; \mathbf{x})$ , is the ratio of the relative change in willingness to pay to the relative change in price from  $q_1$  to  $q_2$ :

$$\nu_A(q_1, q_2; \mathbf{x}) = \frac{\pi(q_2; \mathbf{x}) - \pi(q_1; \mathbf{x})}{\pi(q_1; \mathbf{x})} \times \frac{q_1}{q_2 - q_1},$$

and

$$\nu(q; \mathbf{x}) = \lim_{\tilde{q} \rightarrow q} \nu_A(q, \tilde{q}; \mathbf{x}) = q \cdot \frac{\pi'(q; \mathbf{x})}{\pi(q; \mathbf{x})}.$$

In the absence of any constraints and with profit defined as above (2.1.1), maximising expected profit implies that,

$$\begin{aligned} \pi(q^*; \mathbf{x})V + (q^* - c)\pi'(q^*; \mathbf{x})V &= 0, \\ \implies \frac{\pi'(q^*; \mathbf{x})}{\pi(q^*; \mathbf{x})} &= -\frac{1}{q^* - c}, \end{aligned} \tag{2.1.3}$$

$$\implies \nu(q^*; \mathbf{x}) = -\frac{q^*}{q^* - c}. \tag{2.1.4}$$

(2.1.4) is obtained from (2.1.3) by multiplying by  $q$ , and this result is known as the *price-optimality condition* (Talluri and van Ryzin, 2004, p.190). Since willingness to pay is a strictly decreasing function of price, and  $\pi(q; \mathbf{x}), q > 0$ , elasticity must be negative. In the presence of endogeneity, which generally leads to underestimates of price elasticity, optimal prices will be inflated.

Provided that the willingness to pay function is ‘well-behaved’ in the sense that it is twice continuously differentiable and log-concave,  $(\ln \pi(q; \mathbf{x}))'' < 0$ , then expected profit is a smooth unimodal function. Hence the optimal price can be calculated using standard local search algorithms such as gradient ascent (Phillips, 2010, p.19).

In the unconstrained optimisation problem, transactions are separable and price recommendations can be made independently, as one sale does not affect another. However with the introduction of constraints, this is no longer the case. For example with the delivery of heating oil, constraints arise from capacity limitations and vehicle routing of trucks, as well as the compatibility of different fuels on the same truck. In this case the optimisation is dynamic rather than static; price, cost and the willingness to pay function are indexed by time  $t$ , which is assumed to take discrete values. Then prices at each time point,  $q_t$ , are chosen to maximise the expected contribution to profit over a given time horizon. Constraints such as these significantly increase the complexity of the price optimisation problem, however standard solution approaches can be used if the constraints lead to a convex feasible region.

Although optimal prices have been defined with respect to profit here, profit maximisation need not be the sole pricing objective. If the seller wishes to maximise market share, the objective function becomes expected revenue and the profit function is replaced by  $q$  in (2.1.2). Alternatively, profit can be maximised subject to maintaining a minimum market share by introducing a constraint to specify a lower bound for expected revenue.

In customised pricing, where customers are likely to participate in negotiation with sale agents rather than accept the initial offer price, we can define a range of acceptable prices based on expected profit to guide negotiations. For example agents can be provided with an interval  $(q_L, q_U)$  for which the expected profit will remain within a certain percentage of the optimum (Phillips, 2010, p.28). This combines a decentralised and centralised approach to pricing decisions, whereby pricing recom-



mendations are made with respect to overall corporate goals and agents are given the freedom to negotiate the best possible price within the specified bounds.

### 2.1.2 Estimation of willingness to pay

A demand curve represents the total sales a seller expects to make by offering a price for some fixed period of time, whereas the willingness to pay curve,  $\pi(q; \mathbf{x})$ , specifies the seller's probability of making an individual sale. This probability is expressed as a function of price,  $q$ , and any other information related to the sale,  $\mathbf{x}$ .

Customised pricing is effectively a form of market segmentation; by including non-price factors in our estimate of the customer's probability of acceptance, we are implicitly grouping transactions by price sensitivity. For example, suppose estimation of willingness to pay reveals an inverse relationship between probability of sale and order size. Then using order size as a predictor of willingness to pay leads to price discrimination between small and large orders. Depending on the form in which order size enters the willingness to pay function, this market segmentation can be on a discrete or continuous scale.

Thus in using willingness to pay to inform pricing strategies, market segmentation becomes an issue of model selection. The finer the degree of market segmentation a seller can achieve, the higher his maximum profit. A perfect willingness to pay function would include all factors affecting customers' price sensitivity. However in practice most of this information is unobserved by the seller, for example a customer's income. The result is greater uncertainty around the customer's probability of acceptance and expected profit. Latent variables, in the form of random effects or mixture

components, can be used to capture differences in price sensitivity and support higher degrees of segmentation. We return to this idea in Chapters 3 and 4.

Notice that the issues of market segmentation and endogeneity are related; the purpose of market segmentation is to represent heterogeneity within the customer population, whilst endogeneity is caused by agents recognising and using this heterogeneity in their pricing decisions. Since this customer heterogeneity is not directly observed by the researcher, endogeneity serves to confound the relationship between price and demand.

A natural approach to modelling willingness to pay is to assume each customer has a maximum price that they are willing to pay for their order, called a *referral price*, and they accept any price below this level (Talluri and van Ryzin, 2004). Given the covariates of the order, such as purchase history and order size, and supposing for now that price is independent of demand, referral prices are assumed to be independently and identically distributed across the customer population according to some density  $f_R(r; \mathbf{x})$ . Then if a customer is randomly selected from this population and offered a price of  $q$ , their probability of acceptance is given by the following survivor function:

$$\pi(q; \mathbf{x}) = \int_q^\infty f_R(r; \mathbf{x}) dr = 1 - F_R(q; \mathbf{x}) \quad (2.1.5)$$

where the distribution function,  $F_R(q; \mathbf{x})$ , indicates the proportion of customers with a referral price lower than  $q$ . Note that this function incorporates both our preference and competitive uncertainty about the transaction. In other words, random variation around the customer's referral price arises from uncertainty over the prices offered by

competitors as well as how the customer will choose between them.

The problem of predicting how likely a customer is to accept a quote,  $\pi(q; \mathbf{x})$ , is equivalent to the problem of estimating the distribution of referral prices across the customer population,  $f_R(r; \mathbf{x})$ . Willingness to pay is estimated from historical transaction data, which includes at a minimum, the prices offered in previous quotes and the corresponding responses from customers (accept or reject). Other relevant information may include the customer's location, details of previous purchases from the seller, delivery date, order size and any other factors which are expected to affect a customer's price sensitivity.

Numerous forms of the willingness to pay function exist, but by far the most popular and widely used is the logit function with an associated logistic distribution for referral price (Phillips, 2005):

$$\pi(q; \mathbf{x}) = \left(1 + e^{-(\alpha_0 + g(\mathbf{x}; \boldsymbol{\beta}) + \alpha_1 q)}\right)^{-1}.$$

$\alpha_0$ ,  $\alpha_1$  and  $\boldsymbol{\beta}$  denote the coefficients of the model and  $g(\mathbf{x}; \boldsymbol{\beta})$  is a linear combination of the covariates,  $\mathbf{x}$ , and  $\boldsymbol{\beta}$ .

The reverse sigmoidal shape of the logistic function has a number of appealing properties; provided that  $\alpha_1 < 0$  the probability of sale monotonically decreases with price, and the slope of the curve is steepest for prices with a willingness to pay of 0.5. We define *absolute sensitivity* as the slope of the willingness to pay curve,  $\pi'(q; \mathbf{x})$ . Then at prices with a low willingness to pay  $\{q : \pi(q; \mathbf{x}) < 0.01\}$ , where customers are almost guaranteed to accept a price offer, absolute sensitivity will be close to zero.

The same is true at prices with a high willingness to pay  $\{q : \pi(q; \mathbf{x}) > 0.99\}$ , where customers will almost surely reject. Absolute sensitivity is maximised at  $\pi(q; \mathbf{x}) = 0.5$ . Typically, prices in this region are close to the average or market price, and since the seller's competitors can be expected to offer prices within this region, price sensitivity will be high.

The logistic willingness to pay function tends to be favoured for its interpretability, as the linear predictor occurs on the log-odds scale. However the probit model is an alternative which shares the same properties as above, and for all practical purposes the two functions are indistinguishable (Cramer, 2003, p.26):

$$\pi(q; \mathbf{x}) = \Phi^{-1}(\alpha_0 + g(\mathbf{x}; \boldsymbol{\beta}) + \alpha_1 q) .$$

Our preference for the probit function arises from the need to model dependence between the quote and referral price, for which the bivariate normal distribution is both tractable and directly interpretable. The logit function has relatively fatter tails than the probit, which means that price sensitivity at high prices has different behaviour under each model. To see this, we can consider  $\pi'(q; \mathbf{x})/\pi(q; \mathbf{x})$ . For both models, as  $q \rightarrow -\infty$ ,  $\pi'(q; \mathbf{x}) \rightarrow 0$  and  $\pi(q; \mathbf{x}) \rightarrow 1$ , and hence  $\pi'(q; \mathbf{x})/\pi(q; \mathbf{x}) \rightarrow 0$ . In other words, price changes have no effect on willingness to pay when prices are extremely low. However at the other extreme, price sensitivity in the limit differs

between each model. As  $q \rightarrow \infty$ , under the logistic model we have that:

$$\begin{aligned} \lim_{q \rightarrow \infty} \frac{\pi(q; \mathbf{x})'}{\pi(q; \mathbf{x})} &= \lim_{q \rightarrow \infty} \frac{\alpha_1}{(1 + e^{(\alpha_0 + g(\mathbf{x}; \boldsymbol{\beta}) + \alpha_1 q)})}, \\ &= \alpha_1. \end{aligned} \quad \text{if } \alpha_1 < 0$$

And under the probit model,

$$\begin{aligned} \lim_{q \rightarrow \infty} \frac{\pi(q; \mathbf{x})'}{\pi(q; \mathbf{x})} &= \lim_{q \rightarrow \infty} \frac{\alpha_1 \phi(\alpha_0 + g(\mathbf{x}; \boldsymbol{\beta}) + \alpha_1 q)}{\Phi(\alpha_0 + g(\mathbf{x}; \boldsymbol{\beta}) + \alpha_1 q)}, \\ &= \lim_{q \rightarrow \infty} \frac{-\alpha_1^3 q \phi(\alpha_0 + g(\mathbf{x}; \boldsymbol{\beta}) + \alpha_1 q)}{\alpha_1 \phi(\alpha_0 + g(\mathbf{x}; \boldsymbol{\beta}) + \alpha_1 q)}, \\ &= \lim_{q \rightarrow \infty} -\alpha_1^2 q, \\ &= -\infty \end{aligned}$$

where the second equality comes from applying L'Hôpital's rule (Taylor, 1952). While price sensitivity tends to a constant for the logit link, behaviour in the tail is less stable for the probit. Since we are unlikely to observe 'extreme' prices, this should not have any consequences for the probit model fit. As a robustness check, the model is fitted using both link functions and the coefficients are compared for consistency.

### 2.1.3 Updating and monitoring

The above framework for willingness-to-pay assumes that customers are *myopic*, and purchase a product as soon as they are offered a price below their referral price. Hence customers do not behave strategically, for example by delaying purchases in the hopes

of lower future prices. Similarly, the price optimisation problem presented in Section 2.1.1 implies that the seller's pricing decisions can be made independently to those of competitors. Of course, a model which allows strategic behaviour on the part of customers and competitors is more realistic. However, such a demand model makes the pricing problem essentially a strategic game between customers and sellers, and this significantly complicates the evaluation of optimal pricing strategies - often to the extent of intractability (Talluri and van Ryzin, 2004).

The myopic model is therefore favoured for its tractability, despite being less realistic. However the model can be partly justified on empirical grounds - namely, that historical transaction data has embedded in it the effects of competitors' and customers' responses to the firm's pricing strategy. Hence, provided that the behaviour of competitors and customers observed in past demand continues to apply to current transactions, the myopic model will remain a suitable representation of demand.

For this reason, it is necessary for the seller to periodically monitor the performance of the willingness to pay model relative to the customer responses being observed. Under macroeconomic shifts or changes to customer preferences and competitive actions, we can expect changes in the underlying referral price distribution and consequently, expected profit. This would result in a new optimal price for each transaction, and current pricing recommendations would no longer be relevant under the new market conditions.

Monitoring is particularly important early in the adoption of a price optimisation system, as customers may change their purchasing behaviour in response to the seller's new pricing strategy. For example, if the optimisation model recommends

consistently pricing higher for a particular segment, this can have a substantial impact on customer retention in that segment (Agrawal and Ferguson, 2007). Similarly, competitors may begin to price more aggressively, in which case the seller's pricing model will underestimate the price sensitivity of customers.

If predictions of purchasing decisions deviate significantly from those observed, the model will need to be re-estimated using newly recorded transaction data. Typically, more recent observations are weighted more heavily than older transactions. Alternatively, Bayesian updating allows new observations to be used to directly update the values of parameters in the willingness to pay function (Phillips, 2010). Bayesian estimation of the willingness to pay model is demonstrated in Chapter 5.

## 2.2 Endogeneity

Endogeneity is a regular feature of observational studies, as variation in the independent variable(s) cannot be controlled and may not be entirely random. Pricing research, much of which is based on empirical data, is no exception to this phenomenon. In list pricing, endogeneity is typically caused by the seller anticipating and reacting to future changes in demand; for example airlines may decrease fares for a particular flight because they anticipate low demand (Phillips et al., 2012). In customised pricing, endogeneity results from interaction between agents and customers, as the agent can form an impression of the customer's willingness to pay before offering them a price.

In this section we discuss *instrumental variable estimation*, the leading approach to

endogeneity correction. Section 2.2.1 motivates and outlines the methodology whilst Section 2.3 presents an overview of the work in this area related to customised pricing.

### 2.2.1 Instrumental variable estimation

Frequently used in empirical economic research, instrumental variable (IV) estimation is a technique for estimating causal effects in situations where controlled experiments are not feasible. The first application is attributed to Wright (1928) in his estimation of demand and supply elasticities of flaxseed oil and today the technique is used in a variety of disciplines from epidemiology to psychology. The canonical example is a simple linear model,

$$y = \beta x + \varepsilon; \quad \varepsilon \sim N(0, \sigma^2), \quad (2.2.1)$$

where our interest lies in estimating the causal relationship between  $y$  and  $x$ . The intercept has been excluded for brevity. The ordinary least squares estimator of  $\beta$ ,

$$\hat{\beta}^{OLS} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}} = \beta + \frac{\mathbf{x}'\boldsymbol{\varepsilon}}{\mathbf{x}'\mathbf{x}},$$

is consistent provided that  $\text{Cov}(x, \varepsilon) = 0$ . However if  $x$  is correlated with the unobserved noise term  $\varepsilon$ , then  $\hat{\beta}^{OLS}$  will be biased and inconsistent, and  $x$  is referred to as an *endogenous regressor*. In this case the OLS estimator captures both the direct effect of  $x$  on  $y$  and the indirect effect caused by the joint dependence of  $x$  and  $y$  on  $\varepsilon$ ; whereas we require a consistent estimate of just the direct effect,  $\beta$ .

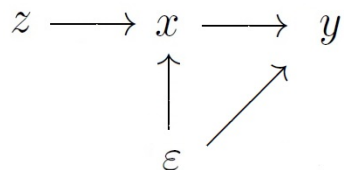


Wooldridge (2002) presents three situations in which  $x$  is found to be endogenous in the estimated model,  $y = \beta x + \varepsilon$ :

1. With omitted variable bias, the true model includes one or more unobserved factors or covariates:  $y = \beta x + \tilde{\beta}\tilde{x} + \tilde{\varepsilon}$ , where  $\tilde{x}$  is unobserved. In the fitted model  $\tilde{x}$  is therefore absorbed by the residual  $\varepsilon$ , which induces a correlation between  $x$  and  $\varepsilon$  if  $x$  is also correlated with  $\tilde{x}$ .
2. Suppose the true model is  $y = \tilde{\beta}\tilde{x} + \tilde{\varepsilon}$  but  $\tilde{x}$  is observed with some measurement error,  $x = \tilde{x} + u$ . The estimated model uses  $x$  as a proxy for  $\tilde{x}$ , and hence  $\varepsilon = \tilde{\varepsilon} - \beta u$ . Then  $x$  and  $\varepsilon$  are correlated as both terms depend on the measurement error,  $u$ .
3. Simultaneity arises when two variables are jointly determined. For example, suppose that the equilibrium relationship between labour demand  $y_d$ , and labour supply  $y_s$ , is described by  $y_d = \alpha y_s + \beta x + \varepsilon_d$  and  $y_s = \tilde{\alpha} y_d + \tilde{\beta} x + \varepsilon_s$ , where  $x$  is the equilibrium wage rate. Estimating either equation results in endogeneity bias, since  $y_d$  will be correlated with  $\varepsilon_s$  and  $y_s$  will be correlated with  $\varepsilon_d$ . This is easy to see by substituting either equation into the other.

Within customised pricing, the endogeneity of price is generally caused by omitted variable bias, as the price offered to a customer is based on unobserved variables which are correlated with their willingness to pay. However the sales process can also be represented by simultaneous equations, since the prices offered by agents are partly determined by a customer's willingness to pay. The model we present in Chapter 3 can be interpreted both in terms of omitted variable bias and simultaneity.

Instrumental variables are introduced to consistently estimate the direct effect of an endogeneous regressor  $x$ , on a quantity of interest  $y$ . A variable  $z$  is said to be an *instrument* if it is correlated with  $x$  but uncorrelated with the error term,  $\varepsilon$ . This leads to the following DAG:



Suppose that  $x$  can be written as a function of  $z$ ,

$$x = \alpha z + u, \quad (2.2.2)$$

where the distribution of  $u$  is unspecified. Substitution of (2.2.2) into (2.2.1) gives:

$$\begin{aligned}
 y &= \beta(\alpha z + u) + \varepsilon, \\
 &= \alpha\beta z + \beta u + \varepsilon, \\
 &= \gamma z + (\beta u + \varepsilon).
 \end{aligned}$$

Hence  $\beta = \gamma/\alpha$  and the instrumental variables estimator of  $\beta$  is defined by:

$$\begin{aligned}
 \hat{\beta}^{IV} &= \frac{\mathbf{z}'\mathbf{y}}{\mathbf{z}'\mathbf{z}} \left( \frac{\mathbf{z}'\mathbf{x}}{\mathbf{z}'\mathbf{z}} \right)^{-1}, \\
 &= \frac{\mathbf{z}'\mathbf{y}}{\mathbf{z}'\mathbf{x}}, \\
 &= \beta + \frac{\mathbf{z}'\boldsymbol{\varepsilon}}{\mathbf{z}'\mathbf{x}}.
 \end{aligned} \quad (2.2.3)$$

The last equality comes from substitution of (2.2.1) for  $\mathbf{y}$  (Bowden and Turkington, 1990). The estimator conditions on  $z$  to capture variation in  $y$  caused directly by  $x$ . The instrument injects variation into  $x$  which is independent of  $\varepsilon$ , thereby allowing us to identify changes in  $y$  caused by  $x$ . We have that  $\hat{\beta}^{IV}$  is a consistent estimator of  $\beta$  under two conditions (Cameron and Trivedi, 2005):

1.

$$\text{plim} \frac{\mathbf{z}'\boldsymbol{\varepsilon}}{N} = \lim_{N \rightarrow \infty} \Pr \left( \left| \frac{1}{N} \sum_{i=1}^N z_i \varepsilon_i \right| \geq \epsilon \right) = 0, \quad (2.2.4)$$

2.

$$\text{plim} \frac{\mathbf{z}'\mathbf{x}}{N} = \lim_{N \rightarrow \infty} \Pr \left( \left| \frac{1}{N} \sum_{i=1}^N z_i x_i \right| \geq \epsilon \right) \neq 0, \quad (2.2.5)$$

for all  $\epsilon > 0$  and a sample size of  $N$ . The first condition is related to independence between  $z$  and  $y$ , and stipulates that the instrument has no direct effect on the response,  $\text{Cov}(z, \varepsilon) = 0$ . The second condition is referred to as *relevance* and requires some association between the instrument and the endogenous regressor,  $\text{Cov}(z, x) \neq 0$ .

Instrumental variable estimation readily extends to the general linear model with multiple regressors,

$$y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon; \quad \varepsilon \sim N(0, \sigma^2).$$

Suppose  $\mathbf{x}$  can be partitioned into  $\mathbf{x} = [\mathbf{x}'_1, \mathbf{x}'_2]$ , where  $\mathbf{x}'_1$  contains endogenous regressors and  $\mathbf{x}'_2$  contains exogenous regressors. Then a valid instrument is  $\mathbf{z} = [\mathbf{z}'_1, \mathbf{x}'_2]$ , where  $\mathbf{x}'_2$  can act as an instrument for itself, since it remains independent of  $\varepsilon$ . The parameters of the model are identifiable provided that the number of instruments is equal to or greater than the number of endogenous regressors,  $\dim(\mathbf{z}_1) \geq \dim(\mathbf{x}_1)$

(Wooldridge, 2002). When  $\dim(\mathbf{z}_1) = \dim(\mathbf{x}_1)$ , a multivariate extension of (2.2.3) is obtained by replacing  $\mathbf{x}$  and  $\mathbf{z}$  with  $\mathbf{X}$  and  $\mathbf{Z}$ :

$$\hat{\beta}^{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}, \quad (2.2.6)$$

with similar generalisations of (2.2.4) and (2.2.5).

In the over-identified case, where the number of instruments is greater than the number of endogenous regressors,  $\dim(\mathbf{z}_1) > \dim(\mathbf{x}_1)$ , a two-stage least squares (2SLS) estimator is used and the instruments are taken as  $\hat{\mathbf{X}} = \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ . The name of the estimator arises from the result that it can be obtained by two consecutive OLS regressions: firstly by regressing  $\mathbf{x}$  against  $\mathbf{z}$  to obtain  $\hat{\mathbf{x}}$ , and secondly by regressing  $\mathbf{y}$  on  $\hat{\mathbf{x}}$ , which returns the IV estimator in (2.2.6) (Theil, 1961).

Although IV estimation can lead to consistent estimation compared to OLS, it also leads to a loss of precision, as can be seen from the variance of the estimator defined in (2.2.3):

$$\hat{\beta}^{IV} \sim N\left(\beta, \sigma^2(\mathbf{x}'\mathbf{z})^{-1}\mathbf{z}'\mathbf{z}(\mathbf{z}'\mathbf{x})^{-1}\right).$$

Small sample properties of the estimator cannot be established analytically. However the asymptotic distribution can be deduced as a special case of the general theory of generalised method of moment estimators (Wooldridge, 2002). The variance can be

expressed as follows:

$$\begin{aligned} V\left(\hat{\beta}^{IV}\right) &= \sigma^2 (\mathbf{x}'\mathbf{x})^{-1} \left( \frac{(\mathbf{x}'\mathbf{z})(\mathbf{z}'\mathbf{x})}{(\mathbf{z}'\mathbf{z})(\mathbf{x}'\mathbf{x})} \right)^{-1}, \\ &= \frac{V\left(\hat{\beta}^{OLS}\right)}{r_{x,z}^2}. \end{aligned}$$

$r_{x,z}^2$  is the squared correlation between the endogeneous regressor and instrumental variable. Since  $0 < r_{x,z}^2 < 1$  by definition, we have that  $V\left(\hat{\beta}^{OLS}\right) < V\left(\hat{\beta}^{IV}\right)$ . IV estimators may not always perform better than OLS if the correlation between the instrument and the endogenous variable is weak, since the loss of precision will be greater and standard errors increase. IV estimators can suffer from small sample bias, since  $\mathbf{z}'\varepsilon$  need not be zero; and the bias can be especially pronounced when instruments are weak, as  $\mathbf{z}'\mathbf{x}$  will be small (Bound et al., 1995).

The validity of potential instruments is tested using conditions (2.2.4) and (2.2.5) and can indicate when instruments will perform poorly. For just-identified models, where the number of instruments and endogenous regressors is equal it is not possible to test for the independence of instruments, and their use must be justified with a priori arguments. However for the over-identified case, one such test is the Sargan test, which is based on the observation that the 2SLS residuals should be uncorrelated with the set of instruments if they are truly independent (Sargan, 1976). The 2SLS residuals are regressed against the instrumental variables and the resulting coefficient of determination is used to test the null hypothesis that the instruments are uncorrelated with  $\varepsilon$ .

The strength of the instruments can be assessed by the F-statistic and coefficient of partial determination, which measure the marginal contribution of one explanatory variable when all others are already included in the model. A low value for either of these statistics in the regression of  $x$  onto  $z$  suggests a weak association between the instrument and endogenous regressor. There is empirical and theoretical evidence that IV estimation with weak instruments is less efficient than OLS, and hence their inclusion in a model must be well supported to obtain stable and accurate predictions.

When the endogenous regressor is temporal, such as annual inflation, a popular choice of instrument is a lagged version of the regressor. For example, the previous year's inflation rate is an effective instrument for current inflation as the rate of inflation is highly autocorrelated; however past inflation provides little predictive power for other macroeconomic variables such as unemployment, after having accounted for current inflation. More generally, it can be difficult to find variables which satisfy the above conditions. For demand curve estimation, a suitable instrument for price must be correlated with price itself but have no direct effect on the quantity demanded. One candidate is a variable that effects market supply, since this also affects price, but is not a direct determinant of demand.

## 2.3 Review of literature on the use of IV estimation in pricing

Pricing is a well researched subject within several academic fields including operational research, marketing, management science and economics. Yet only recently has

the topic received attention within the context of customised pricing environments. Business-to-business (B2B) trade falls under this category and is estimated to generate half of US transactions (Stein, 2013). The importance of customised pricing in the economy would suggest there are substantial gains to improved pricing decisions within this area. However, owing to the seller's freedom to vary prices, these businesses often have complex pricing processes which are difficult to analyse statistically.

In their exploratory analysis of B2B price changes, Elmaghraby et al. (2012) find evidence that prices do not move in the direction or magnitude indicated by cost changes. They recognise that sales agents' estimates of referral price distributions may differ from theoretical models of buyer behaviour. Agrawal and Ferguson (2007) estimate and compare simple willingness to pay functions for B2B data, which include different degrees of segmentation based on order size. The models were evaluated by the improvement in expected profits, after optimising prices for each transaction. They observe that the model providing better goodness of fit to the data also results in higher expected profit improvements.

Underestimation of price sensitivity has been reported by numerous authors. In a meta-analysis of the determinants of price elasticity using 81 studies over the period 1961-2004, Bijmolt et al. (2005) found that elasticity estimates from studies which accounted for endogeneity exceeded estimates from studies with no correction by an average factor of 0.5. The analysis excluded elasticities derived from business-to-business markets, for which the difference is expected to be much higher. Similar results have been found elsewhere; estimating a logit willingness to pay function for Marriott's 'Group Price Optimizer', Hormby et al. (2010) comment that the fitted

curves implied price-insensitive demand. As a result, profit-maximising prices returned by the optimiser were unreasonably high.

In another study of demand for basic and premium cable packages by Petrin and Train (2010), price is expected to be correlated with unobserved attributes, such as quality of programming. Without correction, aggregate demand is estimated to increase with price and negative price elasticity is recovered after accounting for the unobserved attributes. Phillips et al. (2012) sets out to compare the presence of endogeneity in datasets from two auto-lenders. The first is an on-line lender, with no face-to-face interactions and therefore unrecorded attributes, whilst the second is a direct lender in which the sales process involves face-to-face interactions. Hence in the second case price can be based on characteristics of the borrower which are not recorded in the data. No evidence of endogeneity is found for the on-line data set, whereas for the direct lender, price elasticity estimates increase by a factor of 1.2 following correction for endogeneity.

Both of the previously mentioned papers employ *Hausman-type* instruments for endogeneity correction. The term originates from a study of price elasticities for retail goods (Hausman et al., 1994), in which the average price of a product in other markets is used as an instrument for price in a given market. Other markets can refer to different geographical regions, which share the same marginal cost characteristics and will therefore offer similar prices. However, prices in other areas will have low association with local buyer behaviour as they are averaged over the customer population, and will not reflect local demand shocks such as advertising. Although popular, Hausman instruments have become controversial as independence from local demand



may be too difficult to defend. Their use in consumer choice models (Train, 2009) requires instruments to vary across brands in the market, but cross-brand variation can be low in practice.

A widely used approach to endogeneity correction is the BLP method which is used to estimate demand curves in differentiated product markets (Berry et al., 1995). Here the researcher observes market shares for different brands in several markets. A discrete choice model is used to represent a customer's probability of purchasing each brand, whereby the probability of purchase is defined by a multinomial logit function of *utility*. The customer's utility for each brand is based on price and a product characteristic, with some noise, and the customer is assumed to purchase the brand which maximises his utility. The product characteristic,  $\xi$ , affects price but is unobserved by the researcher. The customer's probability of purchasing each brand can be aggregated across all consumers to predict market share. Berry et al. (1995) suggested a two-step approach in which the observed market shares are firstly equated to predicted market shares to estimate the mean utilities for each product. Instrumental variables are introduced which are independent of the unobserved product characteristic,  $E(\xi Z) = 0$ . Since  $\xi$  can be expressed in terms of utility and model parameters, the second step is to solve for the parameters which minimise this sample moment condition. Petrin and Train (2010) criticise the method for being sensitive to sampling error in market shares, and inconsistent when the number of purchase observations per product is small relative to number of customers.

Instead, Petrin and Train (2010) and Phillips et al. (2012) use a control function approach to treat endogeneity, which was first introduced in Rivers and Vuong

(1988). The method is similar to two stage least squares, except that it can be used in non-linear settings such as probit or logit models. Here, the endogenous variable (typically price) is regressed against independent covariates and the residual from this regression is entered as an additional explanatory variable in utility. It is assumed that the endogenous variable can be decomposed into two components: a function of independent covariates and an unobserved component. Correlation between this unobserved component and willingness to pay is then the source of endogeneity. The residuals from the regression of price against covariates are an estimate of the unobserved component, and their inclusion in the utility function ‘controls’ for endogeneity between price and willingness to pay. Since estimation is performed in two stages, and the residuals are an estimate of the unobserved component of price, this introduces an additional error component into the model. Petrin and Train (2003) propose a correction for the standard errors of coefficients to account for this extra variation, and this adapted for our model in Section 3.4.1.

One of the few studies to address endogeneity in customised pricing is Zhang et al. (2014) who include wholesale cost as an instrument for price, in their Bayesian model of purchasing behaviour. They argue that prices can have long term effects in B2B markets because of the importance of buyer-seller relationships, and select prices to maximise long term profitability. Customers make multiple, interrelated decisions for each transaction: when to buy, how much to buy, whether to request a quote or order directly without a quote, and whether to accept the quote. They use a multivariate hidden Markov model to represent these purchasing decisions and define a latent state for each buyer (price sensitive or price insensitive) which governs their

behaviour over time. The results provide little indication of endogeneity bias, as the introduction of an instrumental variable produces only a marginal improvement in model performance.

## 2.4 Bayesian estimation

Suppose we have a model for some variable of interest,  $Y \in \mathcal{Y}$ , which is parameterised by  $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ . The likelihood function,  $p(\mathbf{y}|\boldsymbol{\theta})$ , returns the joint probability density of a sample under this model, for a given set of parameter values. In Bayesian estimation, model inference is performed on a posterior distribution which combines prior knowledge of the model parameters with the the likelihood function. Prior knowledge of the parameters is represented by a probability distribution for  $\boldsymbol{\theta}$ ,  $p(\boldsymbol{\theta})$ , and is specified before any data is observed. The posterior gives the probability distribution of the parameters, conditional on the observed data and follows from Bayes' theorem:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{p(\mathbf{y})}.$$

The marginal likelihood,  $p(\mathbf{y})$ , measures the probability of the data under a given model, after marginalising over the parameters:

$$p(\mathbf{y}) = \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})d\boldsymbol{\theta}.$$

Typically this function does not exist in closed form, and the posterior is only known up to a normalising constant,  $p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})$ . From the posterior, we can learn

the expected value of  $\boldsymbol{\theta}$  based on prior knowledge and information from the data:

$$\mathbb{E}(\boldsymbol{\theta}|\mathbf{y}) = \int_{\boldsymbol{\Theta}} \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y}) \, \mathrm{d}\boldsymbol{\theta}. \quad (2.4.1)$$

Our uncertainty in  $\boldsymbol{\theta}$  is represented by a credible interval, which is defined as follows: the  $(1 - \alpha)\%$  interval  $(\boldsymbol{\theta}_l, \boldsymbol{\theta}_u)$  contains the true parameter value with probability  $(1 - \alpha)$ . Hence,

$$\int_{(\boldsymbol{\theta}_l, \boldsymbol{\theta}_u)} p(\boldsymbol{\theta}|\mathbf{y}) \, \mathrm{d}\boldsymbol{\theta} = 1 - \alpha. \quad (2.4.2)$$

Notice that for a given value of  $\alpha$ , the interval is not uniquely defined. The narrowest credible interval gives the highest posterior density (HPD) interval, and any point within this interval has a higher posterior density than other point outside of the interval.

Predictions for new observations are made from the posterior predictive distribution,

$$p(y^*|\mathbf{y}) = \int_{\boldsymbol{\Theta}} p(y^*|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}) \, \mathrm{d}\boldsymbol{\theta}. \quad (2.4.3)$$

This is the expectation of the likelihood of a new observation,  $p(y^*|\boldsymbol{\theta})$ , over the posterior distribution. In other words, this is the probability distribution for a new observation  $y^*$ , given the data already observed, after averaging over any uncertainty in the parameters. The expectation and variance of  $Y^*$  with respect to this distribution are used to form predictions and prediction intervals for  $Y^*$ .

For most models of practical use, the posterior distribution is analytically intractable and quantities such as (2.4.1) cannot be calculated directly. In this case, simulation methods are employed to sample from the posterior and the resulting samples are used for model inference. Suppose  $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(M)}$  is one such sample. Then a Monte Carlo estimate of (2.4.1), the posterior expectation of  $\boldsymbol{\theta}$  is given by:

$$\hat{\mathbb{E}}(\boldsymbol{\theta}|\mathbf{y}) = \frac{1}{M} \sum_{m=1}^M \boldsymbol{\theta}^{(m)}.$$

Likewise, for given value of  $y^*$ , the posterior predictive density is estimated by:

$$\hat{p}(y^*|\mathbf{y}) = \frac{1}{M} \sum_{m=1}^M p(y^*|\boldsymbol{\theta}^{(m)}).$$

And credible and HPD intervals are estimated by the sample quantiles of  $\{\boldsymbol{\theta}^{(m)}\}_{m=1}^M$ .

Markov chain Monte Carlo (MCMC) describes a set of methods for sampling from the posterior distribution. The idea is to formulate a Markov chain on the parameter space which has the posterior as a stationary distribution (Rossi et al., 2005). Due to the Markov property, the current state of the chain,  $\boldsymbol{\theta}^{(m)}$ , depends only on the previous state,  $\boldsymbol{\theta}^{(m-1)}$ , and the transition kernel defines the conditional distribution of  $\boldsymbol{\theta}^{(m)}|\boldsymbol{\theta}^{(m-1)}$ . The chain is initialised at some value,  $\boldsymbol{\theta}^{(0)} \in \boldsymbol{\Theta}$ , and subsequent values of the chain are drawn from a proposal distribution,  $q(\boldsymbol{\theta}^{(m)}|\boldsymbol{\theta}^{(m-1)})$ , based on the transitional kernel.

One of the most general MCMC schemes is the Metropolis-Hastings (MH) algorithm (e.g. Robert and Casella (2005)). Here, given that the current state of the chain is  $\boldsymbol{\theta}^{(m)}$ , the chain is updated as follows:

1. Draw  $\boldsymbol{\theta}^*$  from the density  $q(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(m)})$ .

2. Calculate the acceptance probability,

$$\alpha(\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^*) = \min \left( 1, \frac{p(\boldsymbol{\theta}^*) p(\mathbf{y}|\boldsymbol{\theta}^*) q(\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(m)}) p(\mathbf{y}|\boldsymbol{\theta}^{(m)}) q(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(m)})} \right).$$

3. Draw  $u \sim U(0, 1)$ .

4. Set

$$\boldsymbol{\theta}^{(m+1)} = \begin{cases} \boldsymbol{\theta}^* & \text{if } u \leq \alpha(\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^*), \\ \boldsymbol{\theta}^{(m)} & \text{otherwise.} \end{cases}$$

Notice that the acceptance probability excludes the normalising constant,  $p(\mathbf{y})$ , from the posterior density, since this is cancelled out in the fraction. The random walk Metropolis (RWM) is a special case of the MH algorithm in which the proposal distribution is symmetric and centred on the current state of the chain. Hence  $q(\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^*) = q(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(m)})$ , and the  $q(\cdot)$  terms are cancelled out in the acceptance probability.

The proposal distribution is crucial to the quality of the posterior sample. If the proposal is a poor representation of the target distribution, in this case the posterior, the values of the chain will be highly correlated and convergence to the target will be slow. An alternative approach, Gibbs sampling, is available when  $\boldsymbol{\theta}$  can be partitioned into  $K$  components,  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K\}$ , and each component has a conditional posterior distribution of known form. That is,  $p(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{-k}, \mathbf{y})$  ( $k = 1, \dots, K$ ) is easy to sample

from, where  $\boldsymbol{\theta}_{-k} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{k-1}, \boldsymbol{\theta}_{k+1}, \dots, \boldsymbol{\theta}_K\}$ . Given an initial set of values,  $\boldsymbol{\theta}^{(0)} = \{\boldsymbol{\theta}_1^{(0)}, \dots, \boldsymbol{\theta}_K^{(0)}\}$ , the Gibbs sampler updates each  $\boldsymbol{\theta}_k$ , conditional on the values of the other  $(K - 1)$  components.

1. Draw  $\boldsymbol{\theta}_1^{(m+1)}$  from  $p(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2^{(m)}, \dots, \boldsymbol{\theta}_K^{(m)}, \mathbf{y})$ .
2. Draw  $\boldsymbol{\theta}_2^{(m+1)}$  from  $p(\boldsymbol{\theta}_2 | \boldsymbol{\theta}_1^{(m+1)}, \boldsymbol{\theta}_3^{(m)}, \dots, \boldsymbol{\theta}_K^{(m)}, \mathbf{y})$ .
- $\vdots$
- K. Draw  $\boldsymbol{\theta}_K^{(m+1)}$  from  $p(\boldsymbol{\theta}_K | \boldsymbol{\theta}_2^{(m+1)}, \dots, \boldsymbol{\theta}_{K-1}^{(m+1)}, \mathbf{y})$ .

Repeatedly iterating over the  $K$  components leads to a valid MCMC scheme and provides a sample from the joint posterior.

When one or more of the conditional posteriors are not a known distribution, a RWM or MH step can be used to update the component. For example, suppose that  $p(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_{2:K}, \mathbf{y})$  cannot be sampled from directly. Then a proposal distribution for  $\boldsymbol{\theta}_1$  is defined,  $q(\boldsymbol{\theta}_1^{(m)} | \boldsymbol{\theta}_1^{(m-1)})$ . In each iteration of the Gibbs sampling scheme, a jump is proposed from  $q$  and accepted or rejected in the usual way. However, the full posterior in the acceptance probability is replaced by the conditional posterior,

$$\alpha(\boldsymbol{\theta}_1^{(m)}, \boldsymbol{\theta}_1^*) = \min \left( 1, \frac{p(\boldsymbol{\theta}_1^*) p(\mathbf{y} | \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_{2:K}^{(m)}) q(\boldsymbol{\theta}_1^{(m)}, \boldsymbol{\theta}_1^*)}{p(\boldsymbol{\theta}_1^{(m)}) p(\mathbf{y} | \boldsymbol{\theta}_1^{(m)}) q(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_1^{(m)})} \right).$$

$\boldsymbol{\theta}_1^{(m)}$  is then updated and steps 2 to K remain unchanged. A mixture of the Metropolis-Hastings and Gibbs sampler in this way is known as Metropolis within Gibbs.

The efficiency of an MCMC scheme, or the number of iterations required for convergence to the stationary distribution, is largely driven by the variance of the proposal.

The variance determines the size of proposal jumps,  $\left|\boldsymbol{\theta}^* - \boldsymbol{\theta}^{(m)}\right|$ , and large jumps are more likely to be rejected, whilst small jumps are slow to traverse the parameter space. In both cases the result is poor mixing, or high autocorrelation in the chain values. When MCMC is performed on a high-dimensional target distribution, suitable tuning of the proposal to avoid these scenarios is difficult to achieve. Instead we can turn to adaptive MCMC, which learns the shape of the target distribution and scales the proposal distribution accordingly. Further details of this algorithm are provided in Section 5.2.3.



# Chapter 3

## Modelling willingness to pay

In this chapter we present a probit model for estimating willingness to pay. Each customer is assumed to have a price threshold for each order, referred to as referral price, such that they accept any quote below this threshold. The context for our model is customised pricing where we can expect some dependence between the price offered to a customer and their referral price, and so a bivariate Gaussian distribution is used to model both variables. Two-stage estimation of the model is described and a procedure for obtaining standard errors is outlined. This is followed by an example with simulated data and a discussion of some possible extensions to the model.

### 3.1 Formulation

Let  $Q$  and  $R$  denote the quote price and associated referral price of the customer at the time of the quote, both of which are assumed to have normal marginal distributions.

The dependence between  $Q$  and  $R$  is modelled via their joint distribution:

$$\begin{pmatrix} Q \\ R \end{pmatrix} \sim N \left( \begin{bmatrix} \mathbf{x}'_q \boldsymbol{\beta}_q \\ \mathbf{x}'_r \boldsymbol{\beta}_r \end{bmatrix}, \begin{bmatrix} \sigma_q^2 & \rho \sigma_q \sigma_r \\ \rho \sigma_q \sigma_r & \sigma_r^2 \end{bmatrix} \right). \quad (3.1.1)$$

The mean of each distribution is a linear combination of covariates:

$$\begin{aligned} \mathbf{x}_r &= \{ \mathbf{x}^{(c)}, \mathbf{x}^{(r)} \}, \quad \boldsymbol{\beta}_r = \{ \boldsymbol{\beta}_r^{(c)}, \boldsymbol{\beta}_r^{(r)} \}, \\ \mathbf{x}_q &= \{ \mathbf{x}^{(c)}, \mathbf{x}^{(q)} \}, \quad \boldsymbol{\beta}_q = \{ \boldsymbol{\beta}_q^{(c)}, \boldsymbol{\beta}_q^{(q)} \}. \end{aligned}$$

Here  $\mathbf{x}^{(c)}$ ,  $\mathbf{x}^{(q)}$  and  $\mathbf{x}^{(r)}$  are recorded covariates that are observable by the sales agent and statistical modeller. The set of common covariates,  $\mathbf{x}^{(c)}$ , affects both the quote price and referral price, whereas  $\mathbf{x}^{(q)}$  is the set of instrumental variables and affects the price offered by the agent, but not willingness to pay. Similarly, covariates in  $\mathbf{x}^{(r)}$  affect referral price only. For ease of notation, define  $\mathbf{x} = \{ \mathbf{x}^{(c)}, \mathbf{x}^{(r)}, \mathbf{x}^{(q)} \}$ , or equivalently  $\mathbf{x} = \mathbf{x}_q \cup \mathbf{x}_r$ . The parameter  $\rho$  represents any residual correlation between quote and referral price which is not captured by observable covariates.

A bivariate Gaussian was chosen to represent the joint dependence between quote price and referral price, firstly because a Gaussian assumption is generally reasonable for quote price or a transformation of quote price (see Section 3.6.1) and secondly, since referral price is a latent variable and only the binary customer responses are observed, more complex dependence structures such as those expressed by a copula are difficult to identify from the data.

Let  $Y$  be a binary variable indicating whether or not the quote was accepted. Then given a quote price of  $q$ , the quote will be accepted if the customer's referral price exceeds  $q$ .

$$Y = \begin{cases} 1 & \text{if } R > q \text{ i.e. the quote is accepted} \\ 0 & \text{otherwise.} \end{cases}$$

Hence conditional on  $q$ ,  $Y$  has a Bernoulli distribution where the probability of a sale is given by the willingness to pay function:  $Y|q \sim \text{Bernoulli}(\pi(q; \mathbf{x}, \boldsymbol{\theta}))$ . In order to find  $\pi(q; \mathbf{x}, \boldsymbol{\theta})$ , we require the conditional distribution of referral price (e.g. Gut (2009, p.127)):

$$R|Q = q \sim N \left( \mathbf{x}'_r \boldsymbol{\beta}_r + \rho \frac{\sigma_r}{\sigma_q} (q - \mathbf{x}'_q \boldsymbol{\beta}_q) , \sigma_r^2 (1 - \rho^2) \right). \quad (3.1.2)$$

Denote the mean of this distribution by  $\mu_{r|q}$  and the parameters of the model by  $\boldsymbol{\theta} = \left\{ \boldsymbol{\beta}_q^{(c)}, \boldsymbol{\beta}_q^{(q)}, \sigma_q, \boldsymbol{\beta}_r^{(c)}, \boldsymbol{\beta}_r^{(r)}, \sigma_r, \rho \right\}$ . Although the customer was offered a price of  $q$  by the agent, we can imagine a new hypothetical price  $\tilde{q}$ . The probability that the customer's referral price exceeds this hypothetical price, given that they were offered a price of  $q$  is given by:

$$\begin{aligned} P(R > \tilde{q} | q, \mathbf{x}) &= 1 - \Phi \left( \frac{\tilde{q} - \mu_{r|q}}{\sigma_r \sqrt{1 - \rho^2}} \right) \\ &= \Phi \left( \frac{\mathbf{x}'_r \boldsymbol{\beta}_r + \rho \frac{\sigma_r}{\sigma_q} (q - \mathbf{x}'_q \boldsymbol{\beta}_q) - \tilde{q}}{\sigma_r \sqrt{1 - \rho^2}} \right) \\ &= \Phi \left( \frac{1}{\sqrt{1 - \rho^2}} \left[ \frac{1}{\sigma_r} \mathbf{x}'_r \boldsymbol{\beta}_r - \frac{1}{\sigma_r} \tilde{q} + \frac{\rho}{\sigma_q} q - \frac{\rho}{\sigma_q} \mathbf{x}'_q \boldsymbol{\beta}_q \right] \right). \end{aligned} \quad (3.1.3)$$

Let the willingness to pay function, or the probability of the customer accepting

the sales agent's offer of  $q$ , is given by:

$$\begin{aligned}
 \pi(q; \mathbf{x}, \boldsymbol{\theta}) &= P(Y = 1 | q, \mathbf{x}) \\
 &= P(R > q | q, \mathbf{x}) \\
 &= \Phi \left( \frac{1}{\sqrt{1 - \rho^2}} \left[ \frac{1}{\sigma_r} \mathbf{x}'_r \boldsymbol{\beta}_r - q \left( \frac{1}{\sigma_r} - \frac{\rho}{\sigma_q} \right) - \frac{\rho}{\sigma_q} \mathbf{x}'_q \boldsymbol{\beta}_q \right] \right). \quad (3.1.4)
 \end{aligned}$$

Now consider a new customer, for whom there is no existing quote. In this case the willingness to pay function is obtained directly from the marginal referral price distribution, as the price offered to the customer is a decision variable and independent of  $R$ . That is,  $P(R > \tilde{q} | q, \mathbf{x})$  is replaced by  $P(R > \tilde{q} | \mathbf{x}_r)$ , since the agent has not made a price offer. Instead, referral price is only affected by the covariates  $\mathbf{x}_r$ , and the probability of the customer accepting a price of  $\tilde{q}$  is given by:

$$\pi(\tilde{q}; \mathbf{x}_r, \boldsymbol{\theta}) = \Phi \left( \frac{1}{\sigma_r} \mathbf{x}'_r \boldsymbol{\beta}_r - \frac{1}{\sigma_r} \tilde{q} \right). \quad (3.1.5)$$

This is simply (3.1.3) with  $\rho = 0$ . In Chapter 5 we consider willingness to pay for existing customers, under a new hypothetical price of  $\tilde{q}$ . In this case we have observed the customer's response to the sales agent's offer and can estimate the probability of sale given  $q$  and  $y$ .

## 3.2 Price elasticity

In Section 2.1.1 the idea of price elasticity was introduced as a measure of the responsiveness of willingness to pay to changes in price. Price elasticities are often of interest

to practitioners as they are used to calculate optimal prices, and also to compare price sensitivity over time and across products.

Under endogeneity, price changes have both a direct and indirect effect on willingness to pay and to separate these effects, it is helpful to think of the probability of a sale in terms of (3.1.3). Since the agent's offer price,  $q$ , is correlated with referral price, it enters into the mean of the conditional referral price distribution,  $\mu_{r|q}$ , and affects willingness to pay indirectly (3.1.2). The direct effect of price becomes apparent by considering a hypothetical price  $\tilde{q}$ , whereby quotes are only accepted iff  $R > \tilde{q}$ . This is illustrated in Figure 3.2.1, where the probability of a sale is represented by the shaded area in the upper tail of the referral price distribution.

Given that the customer was quoted a price of  $q$ , a higher hypothetical price signifies a higher quantile of the referral price distribution and hence a lower probability of sale. That is,  $P(R > \tilde{q}_1 | q, \mathbf{x}) > P(R > \tilde{q}_2 | q, \mathbf{x})$  for  $\tilde{q}_1 < \tilde{q}_2$ . In conjunction with this and provided that  $\rho > 0$ , a higher quote price indicates a higher expected referral price, so that  $\mu_{r|q_1} < \mu_{r|q_2}$  when  $q_1 < q_2$ . As a consequence, if two customers have identical covariates and one customer is offered  $q_1$  and a second is offered  $q_2$ , the probability of the first customer's referral price exceeding  $\tilde{q}$  is lower than the same probability for the second customer. That is,  $P(R > \tilde{q} | q_1, \mathbf{x}) < P(R > \tilde{q} | q_2, \mathbf{x})$  for  $q_1 < q_2$ .

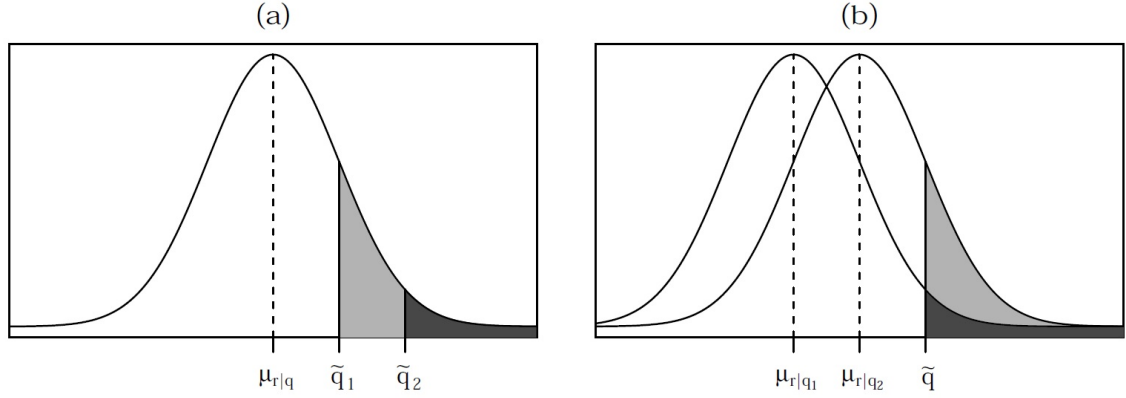


Figure 3.2.1: Direct (a) and indirect (b) effects of price on willingness to pay. The shaded areas represent the probability of sale. (a) Given that the customer was offered a price of  $q$ , an increase in the hypothetical price from  $\tilde{q}_1$  to  $\tilde{q}_2$  causes the probability of sale to decrease from the lighter region to the darker region. (b) If the agent were to offer a quote of  $q_2$  rather than  $q_1$  this would imply a different referral price mean and therefore a shift in the referral price distribution. Hence for a given hypothetical price,  $\tilde{q}$ , the probability of sale is higher under  $q_2$  than  $q_1$ .

The combination of these two effects means that the total effect of price on the probability of the customer accepting the quote,  $P(R > q|q)$ , is not obvious. Therefore in the presence of endogeneity, where quote price also affects the probability of sale indirectly, the proportion of acceptances observed in the data may not decrease with price. In particular, a higher price may in fact be associated with a higher willingness to pay, due to the underlying relationship between  $q$  and expected referral price. For this reason, it is useful to define two types of price elasticity for customised pricing: *naive price elasticity* and *true price elasticity*. Naive price elasticity measures the overall effect of price on willingness to pay, and is obtained by differentiating (3.1.4)

with respect to  $q$ :

$$\begin{aligned}\nu(q; \mathbf{x}, \boldsymbol{\theta}) &= \frac{q}{\pi(q; \mathbf{x}, \boldsymbol{\theta})} \cdot \pi'(q; \mathbf{x}, \boldsymbol{\theta}) \\ &= -\frac{1}{\sqrt{1-\rho^2}} \left( \frac{1}{\sigma_r} - \frac{\rho}{\sigma_q} \right) \frac{q}{\pi(q; \mathbf{x}, \boldsymbol{\theta})} \phi \left( \frac{1}{\sqrt{1-\rho^2}} \left[ \mathbf{x}'_r \frac{\boldsymbol{\beta}_r}{\sigma_r} - q \frac{1}{\sigma_r} + \varepsilon_q \frac{\rho}{\sigma_q} \right] \right).\end{aligned}\tag{3.2.1}$$

Notice that positive price elasticity is observed iff  $\rho > \frac{\sigma_q}{\sigma_r}$ . In other words, if residual correlation is high relative to the ratio of quote and referral price standard deviation. Alternatively, true price elasticity captures only the direct effect of price on willingness to pay (illustrated by (a) in Figure 3.2.1). This obtained by differentiating (3.1.3) with respect to  $\tilde{q}$ :

$$\nu(q; \mathbf{x}, \boldsymbol{\theta}) = -\frac{1}{\sigma_r \sqrt{1-\rho^2}} \frac{q}{\pi(q; \mathbf{x}, \boldsymbol{\theta})} \phi \left( \frac{1}{\sqrt{1-\rho^2}} \left[ \mathbf{x}'_r \frac{\boldsymbol{\beta}_r}{\sigma_r} - q \frac{1}{\sigma_r} + \varepsilon_q \frac{\rho}{\sigma_q} \right] \right).\tag{3.2.2}$$

True price elasticity will always be negative as a higher hypothetical price,  $\tilde{q}$ , has a lower probability of sale, all other things unchanged. However it cannot be observed directly, since only the overall relationship between price and willingness to pay is observed in the sales data.

Having estimated the parameters of the model (by the two-step method to be discussed in Section, 3.3.2), we may wish to offer a quote to a new customer with covariates  $\mathbf{x}_r$ . In this case there is no existing quote price and we can consider an arbitrary price,  $\tilde{q}$ , which is chosen without reference to the customer. Then price elasticity is calculated by differentiating (3.1.5) w.r.t  $\tilde{q}$ :

$$\nu(\tilde{q}; \mathbf{x}_r, \boldsymbol{\theta}) = -\frac{1}{\sigma_r} \frac{\tilde{q}}{\pi(\tilde{q}; \mathbf{x}_r, \boldsymbol{\theta})} \phi \left( \frac{1}{\sigma_r} \mathbf{x}_r' \boldsymbol{\beta}_r - \frac{1}{\sigma_r} \tilde{q} \right).$$

In this instance, price elasticity is negative because the price quoted to the customer is independent of their referral price and has no indirect relationship with willingness to pay.

### 3.3 Parameter estimation

Given a sample of  $n$  quotes with prices  $\mathbf{q}$ , customer responses  $\mathbf{y}$  and covariate matrix  $\mathbf{X}$ , with  $i^{th}$  row  $\mathbf{x}_i = \left\{ \mathbf{x}_i^{(c)}, \mathbf{x}_i^{(r)}, \mathbf{x}_i^{(q)} \right\}$ , the likelihood of  $\boldsymbol{\theta}$  is expressed by:

$$\begin{aligned} L(\boldsymbol{\theta} | \mathbf{y}, \mathbf{q}, \mathbf{X}) &= \prod_{i=1}^n P(y_i | q_i) P(q_i) \\ &= \prod_{i=1}^n \pi(q_i; \mathbf{x}_i, \boldsymbol{\theta})^{y_i} (1 - \pi(q_i; \mathbf{x}_i, \boldsymbol{\theta}))^{1-y_i} \cdot \frac{1}{\sigma_q} \phi \left( \frac{q_i - \mathbf{x}_{q,i}' \boldsymbol{\beta}_q}{\sigma_q} \right). \end{aligned} \quad (3.3.1)$$

Separating  $\boldsymbol{\theta}$  into the quote price parameters,  $\boldsymbol{\theta}_q = \{\boldsymbol{\beta}_q, \sigma_q\}$ , and referral price parameters,  $\boldsymbol{\theta}_r = \{\boldsymbol{\beta}_r, \sigma_r, \rho\}$ , we can express the likelihood as a product of the conditional likelihood for  $\boldsymbol{\theta}_r$  given  $\boldsymbol{\theta}_q$ , and the marginal likelihood of  $\boldsymbol{\theta}_q$ :

$$\begin{aligned} L(\boldsymbol{\theta} | \mathbf{y}, \mathbf{q}, \mathbf{X}) &= \pi(\mathbf{y}; \mathbf{q}, \mathbf{X}, \boldsymbol{\theta}) \cdot \pi(\mathbf{q}; \mathbf{X}, \boldsymbol{\theta}_q) \\ &= \left[ \prod_{i=1}^n \pi(q_i; \mathbf{x}_i, \boldsymbol{\theta})^{y_i} (1 - \pi(q_i; \mathbf{x}_i, \boldsymbol{\theta}))^{1-y_i} \right] \left[ \prod_{i=1}^n \frac{1}{\sigma_q} \phi \left( \frac{q_i - \mathbf{x}_{q,i}' \boldsymbol{\beta}_q}{\sigma_q} \right) \right]. \end{aligned} \quad (3.3.2)$$



### 3.3.1 Identifiability

The quote price parameters can, if necessary, be estimated directly from observed quote prices and covariates, although this does not use the information in  $L(\boldsymbol{\theta}_r|\boldsymbol{\theta}_q)$ .

They are selected to maximise the pseudo-marginal likelihood,  $L(\boldsymbol{\theta}_q|\mathbf{q}, \mathbf{X})$ :

$$\left\{\hat{\boldsymbol{\beta}}_q, \hat{\sigma}_q\right\} = \arg \max_{\boldsymbol{\beta}_q, \sigma_q} \prod_{i=1}^n \frac{1}{\sigma_q} \phi\left(\frac{q_i - \mathbf{x}'_{q,i} \boldsymbol{\beta}_q}{\sigma_q}\right), \quad (3.3.3)$$

which is equivalent to obtaining the OLS estimates in a regression of price on the covariates.

Suppose for now that instrumental variables are not included in the model, so that  $\mathbf{x}^{(q)} = \emptyset$  and  $\mathbf{x}_q = \mathbf{x}^{(c)}$ . For simplicity, also assume  $\mathbf{x}^{(r)} = \emptyset$  and  $\mathbf{x}_r = \mathbf{x}^{(c)}$ . The willingness to pay function then becomes:

$$\pi(q; \mathbf{x}, \boldsymbol{\theta}) = \Phi\left(\frac{1}{\sqrt{1-\rho^2}} \left[ \mathbf{x}^{(c)'} \left( \frac{1}{\sigma_r} \boldsymbol{\beta}_r - \frac{\rho}{\sigma_q} \boldsymbol{\beta}_q \right) - q \left( \frac{1}{\sigma_r} - \frac{\rho}{\sigma_q} \right) \right]\right). \quad (3.3.4)$$

Under this set up, the model is non-identifiable because it is possible for different combinations of referral price parameters to produce identical values of the conditional likelihood. To see this, notice that the two parameter combinations  $\boldsymbol{\theta}^{(1)} = \left\{\hat{\boldsymbol{\beta}}_q, \hat{\sigma}_q, \boldsymbol{\beta}_r^{(1)}, \sigma_r^{(1)}, \rho^{(1)}\right\}$  and  $\boldsymbol{\theta}^{(2)} = \left\{\hat{\boldsymbol{\beta}}_q, \hat{\sigma}_q, \boldsymbol{\beta}_r^{(2)}, \sigma_r^{(2)}, \rho^{(2)}\right\}$  will produce the same values of the willingness to pay function if the following two equations are satisfied:

$$\frac{1}{\sqrt{1-\rho^{(1)2}}} \left( \frac{1}{\sigma_r^{(1)}} \boldsymbol{\beta}_r^{(1)} - \frac{\rho^{(1)}}{\hat{\sigma}_q} \hat{\boldsymbol{\beta}}_q \right) = \frac{1}{\sqrt{1-\rho^{(2)2}}} \left( \frac{1}{\sigma_r^{(2)}} \boldsymbol{\beta}_r^{(2)} - \frac{\rho^{(2)}}{\hat{\sigma}_q} \hat{\boldsymbol{\beta}}_q \right), \quad (3.3.5)$$

$$\frac{1}{\sqrt{1 - \rho^{(1)2}}} \left( \frac{1}{\sigma_r^{(1)}} - \frac{\rho^{(1)}}{\hat{\sigma}_q} \right) = \frac{1}{\sqrt{1 - \rho^{(2)2}}} \left( \frac{1}{\sigma_r^{(2)}} - \frac{\rho^{(2)}}{\hat{\sigma}_q} \right). \quad (3.3.6)$$

Supposing  $\beta_r$  contains  $p_r$  unknown parameters, (3.3.5) and (3.3.6) consist of  $p_r + 1$  equations in  $p_r + 2$  unknowns and therefore admit no unique solution. Given  $\rho^{(1)}$ ,  $\beta_r^{(1)}$  and  $\sigma_r^{(1)}$ , for any  $\sigma_r^{(2)} > 0$ , we could fix  $\rho^{(2)}$  arbitrarily between -1 and 1. Then (3.3.5) provides a linear expression for  $\beta_r^{(2)}$ . However, in general there is a continuum of pairs of values  $(\sigma_r^{(2)}, \rho^{(2)})$  that satisfy (3.3.6). The same is true when referral price covariates,  $\mathbf{x}^{(r)}$ , are introduced.

However when  $\mathbf{x}^{(q)}$  is non-empty and instrumental variables are included in (3.3.4), the model no longer suffers from systematic non-identifiability. This is because another equation is introduced, in addition to (3.3.5) and (3.3.6):

$$\frac{1}{\sqrt{1 - \rho^{(1)2}}} \frac{\rho^{(1)}}{\hat{\sigma}_q} = \frac{1}{\sqrt{1 - \rho^{(2)2}}} \frac{\rho^{(2)}}{\hat{\sigma}_q},$$

which allows  $\rho$  (and therefore  $\beta_r$  and  $\sigma_r$ ) to be identified.

### 3.3.2 Two-stage estimation

There are numerous re-parameterisations of the willingness to pay function (3.1.4), although arguably the most intuitive form is given by:

$$\pi(q; \mathbf{x}, \boldsymbol{\theta}) = \Phi \left( \frac{1}{\sqrt{1 - \rho^2}} \left[ \mathbf{x}_r' \left( \frac{\beta_r}{\sigma_r} \right) - \frac{1}{\sigma_r} q + \frac{\rho}{\sigma_q} \varepsilon_q \right] \right), \quad (3.3.7)$$

which also conforms to the control function approach in Rivers and Vuong (1988). The noise term,  $\varepsilon_q \sim N(0, \sigma_q^2)$ , represents any unexplained variation in price. Then  $q$  can be written as:

$$q = \mathbf{x}_q' \boldsymbol{\beta}_q + \varepsilon_q. \quad (3.3.8)$$

Given that at least one instrument for price is available, the parameters are identifiable and estimates can be obtained in two stages. Firstly,  $\hat{\boldsymbol{\theta}}_q$  is obtained by a straightforward linear regression of  $\mathbf{q}$  on  $\mathbf{x}^{(c)}$  and  $\mathbf{x}^{(q)}$ . The noise terms are estimated by the residuals from this regression:

$$\hat{\varepsilon}_{q,i} = q_i - \mathbf{x}_{q,i}' \hat{\boldsymbol{\beta}}_q. \quad (3.3.9)$$

These are then used alongside price to estimate the referral price parameters in a probit regression:

$$\Phi^{-1}(\pi(q_i; \mathbf{x}, \boldsymbol{\theta})) = \mathbf{x}_i' \boldsymbol{\gamma} + \lambda q_i + \psi \hat{\varepsilon}_{q,i}. \quad (3.3.10)$$

The customer responses  $\mathbf{y}$ , are regressed against the referral price covariates, quote price and quote price residuals to give the corresponding coefficients  $\{\hat{\gamma}, \hat{\lambda}, \hat{\psi}\}$ . These are raw or untransformed versions of the referral price parameters,  $\boldsymbol{\phi}_r$ , and the original parameters are recovered by the following transformations:

$$\rho = \frac{\psi \sigma_q}{(1 + \psi^2 \sigma_q^2)^{1/2}}, \quad (3.3.11)$$

$$\sigma_r = -\frac{(1 + \psi^2 \sigma_q^2)^{1/2}}{\lambda}, \quad (3.3.12)$$

$$\boldsymbol{\beta}_r = -\frac{\boldsymbol{\gamma}}{\lambda}, \quad (3.3.13)$$

The transformations are obtained by solving the equations that emerge from equating (3.3.7) and (3.3.10), and are denoted by  $\boldsymbol{\theta}_r = g(\boldsymbol{\phi}_r)$ .

There are a number of reparameterisations of the willingness to pay function, as  $q$  or  $\varepsilon_q$  can be written in terms of any rearrangement of (3.3.9). For practical purposes, we favour using  $q$  and  $\varepsilon_q$  because there is a one-to-one relationship between the signs of the untransformed and transformed versions of the referral price parameters:  $\text{sgn}(\psi) = \text{sgn}(\rho)$ ,  $\text{sgn}(\lambda) = -\text{sgn}(\sigma_r)$ . In other words, we require a negative coefficient for  $q$  in the probit regression to obtain a valid estimate of the referral price variance, and a positive coefficient for  $\hat{\varepsilon}_q$  to have a sensible estimate of  $\rho$ .

Since the quote price parameters are estimated in the first stage regression despite also featuring in the willingness to pay function, this may imply some loss of information about  $\boldsymbol{\theta}_q$  in the two-stage approach compared to maximisation of the full likelihood (3.3.1). However provided that the sample size is large (as is the case here), the impact on the parameters of interest,  $\boldsymbol{\theta}_r$ , should be small. Typically the quote prices will contain much more information about the quote price parameters than the binary responses of customers, and hence the two-stage approach is warranted. To explore the efficiency loss associated with this assumption, the parameters of the models in the simulated example in Section 3.5 were estimated both by maximising the full likelihood (3.3.2) and by the two stage approach described here. There was a negligible difference between the two sets of parameter estimates.

### 3.3.3 Willingness to pay without endogeneity

In the absence of endogeneity, there are no sources of unobserved dependence between price and the probability of sale and  $\rho = 0$ . The willingness to pay function then simplifies to:

$$\pi(q; \mathbf{x}, \boldsymbol{\theta}) = \Phi \left( \mathbf{x}_r' \left( \frac{\boldsymbol{\beta}_r}{\sigma_r} \right) - \frac{1}{\sigma_r} q \right). \quad (3.3.14)$$

Comparison with (3.3.7) reveals that the endogeneity bias causes the coefficients of the willingness to pay function to inflate by a factor of  $(1 - \rho^2)^{1/2}$ .

When endogeneity correction is used for predictive purposes, that is, to predict the probability of a new quote being accepted, it is enough to normalise the probit coefficients by  $(1 - \rho^2)^{1/2}$ , or equivalently,  $(1 + \psi^2 \sigma_q^2)^{1/2}$ . The new price  $\tilde{q}$ , and covariates  $\tilde{\mathbf{x}}$ , are then entered into the corrected willingness to pay function to predict the probability of a sale.

The focus of this study is historical transaction data, in which prices are not independent of willingness to pay and cannot be changed in isolation. We wish to learn about the dependence between price and willingness to pay, and the degree of endogeneity in the data. For that reason, our interest lies in estimating the parameters of the referral price distribution rather than just the probit coefficients. We discuss the use of the model for prediction in Chapter 5.

## 3.4 Estimation of standard errors

### 3.4.1 Standard errors of parameter estimates

Standard errors for the quote price parameter estimates are obtained in the usual way for a linear regression (e.g. Wooldridge (2002)). The variance of the coefficients is given by  $\text{Var}(\beta_q) = \sigma_q(X_q'X_q)^{-1}$ . Based on the result that  $(n - p_q)s^2/\sigma_q^2 \sim \chi_{n-p_q}^2$ , where  $s^2$  is the variance of the quote price residuals and  $p_q$  is the number of coefficients in the model, a confidence interval for the quote price variance is calculated from the quantiles of a Chi-squared distribution with  $n - p_q$  degrees of freedom (Lehmann and Casella, 2011).

Let  $\hat{\Sigma}_{\phi_r}$  denote the inverse of the observed Fisher Information matrix from the probit regression. Ordinarily, the standard errors of the probit coefficients from (3.3.10) are taken from the diagonal entries of this matrix. However, these standard errors for  $\phi_r$  will not be accurate because the regression includes the residuals from the first stage regression,  $\hat{\varepsilon}_q$ , which introduces an extra source of variation into the equation. Since these residuals are an estimate of the true noise terms for the quote price,  $\varepsilon_q$ , the sampling variance of the estimators for the referral price parameters will be underestimated.

Petrin and Train (2003) outline a bootstrapping procedure for estimating the true standard errors of a two-stage estimator. The first stage regression is repeated with bootstrapped samples of the quote prices and quote price covariates; the quote price residuals are then calculated and the probit regression is repeated by with the new residuals. The variance in the probit coefficients over the bootstrapped samples is

added to the estimated variance of the original probit coefficients,  $\hat{\Sigma}_{\phi_r}$ , to give more accurate standard errors for the referral price parameters.

### Bootstrap procedure

For  $b = 1, \dots, B$ :

Resample the quote price data,  $\mathbf{q}^{(b)}, \mathbf{X}_q^{(b)}$ .

Perform the first stage regression for the new set of prices and calculate the residuals,  $\hat{\epsilon}_q^{(b)}$ :

$$\mathbf{q}^{(b)} = \mathbf{X}_q^{(b)} \hat{\beta}_q^{(b)} + \hat{\epsilon}_q^{(b)}.$$

Perform the probit regression on the customer responses,  $\mathbf{y}$ , using the new set of residuals:

$$\pi(\mathbf{q}; \mathbf{X}) = \Phi \left( \mathbf{X}_r \hat{\gamma}^{(b)} + \hat{\lambda}^{(b)} \mathbf{q} + \hat{\psi}^{(b)} \hat{\epsilon}_q^{(b)} \right).$$

Store the raw coefficients,  $\hat{\phi}_r^{(b)} = \left\{ \hat{\gamma}^{(b)}, \hat{\lambda}^{(b)}, \hat{\psi}^{(b)} \right\}$  and transformed referral price parameters  $\hat{\theta}_r^{(b)} = g \left( \hat{\phi}_r^{(b)} \right)$ .

Denote the sample covariance matrix of the bootstrap estimates for  $\theta_r$  and  $\phi_r$  by  $\Sigma_{\theta_r}^B$  and  $\Sigma_{\phi_r}^B$  respectively.

The adjusted variance of the raw probit coefficients, which accounts for the error in using  $\hat{\epsilon}_q$ , is then given by:

$$\Sigma_{\phi_r}^* = \hat{\Sigma}_{\phi_r} + \Sigma_{\phi_r}^B. \quad (3.4.1)$$

Likewise, the adjusted variance of the referral price parameters is given by:

$$\Sigma_{\theta_r}^* = \mathbf{D}(\hat{\phi}_r)^T \hat{\Sigma}_{\phi_r} \mathbf{D}(\hat{\phi}_r) + \Sigma_{\theta_r}^B. \quad (3.4.2)$$

Here the Delta Method has been applied to the ML variance of the probit coefficients,  $\hat{\Sigma}_{\phi_r}$ , to give the corresponding variance of the transformed parameter estimates,  $\hat{\theta}_r$ . The Delta Method (e.g. Casella and Berger (2002, p.240)) uses a Taylor Series approximation to approximate the mean and variance of a function of a set of random variables:

$$\text{Var}(h(\mathbf{z})) = \nabla h(\mathbf{z})^T \text{Var}(\mathbf{z}) \nabla h(\mathbf{z}).$$

Recall that  $\theta_r = g(\phi_r)$  denotes the set of transformations given in (3.3.11) - (3.3.13). Then  $\mathbf{D}(\hat{\phi}_r)$  is the matrix of partial derivatives of  $\theta_r$  with respect to  $\phi_r$ , evaluated at the MLEs. Given  $p_r$  parameters in  $\beta_r$ ,  $\mathbf{D}(\phi_r)$  can be expressed as follows:

$$\mathbf{D}(\phi_r) = \begin{pmatrix} \frac{\partial \beta_r}{\partial \gamma} & \frac{\partial \sigma_r}{\partial \gamma} & \frac{\partial \rho}{\partial \gamma} \\ \frac{\partial \beta_r}{\partial \lambda} & \frac{\partial \sigma_r}{\partial \lambda} & \frac{\partial \rho}{\partial \lambda} \\ \frac{\partial \beta_r}{\partial \psi} & \frac{\partial \sigma_r}{\partial \psi} & \frac{\partial \rho}{\partial \psi} \end{pmatrix}$$



where,

$$\begin{aligned}
\frac{\partial \beta_{r,j}}{\partial \gamma_j} &= -\frac{1}{\lambda}, & \text{for } j = 1, \dots, p_r \\
\frac{\partial \beta_{r,j}}{\partial \lambda} &= -\frac{\gamma_j}{\lambda^2}, & \text{for } j = 1, \dots, p_r \\
\frac{\partial \sigma_r}{\partial \lambda} &= \frac{1}{\lambda^2} (1 + \psi^2 \sigma_q^2)^{1/2}, \\
\frac{\partial \sigma_r}{\partial \psi} &= -\frac{\psi}{\lambda} \sigma_q^2 (1 + \psi^2 \sigma_q^2)^{-1/2}, \\
\frac{\partial \rho}{\partial \psi} &= \sigma_q (1 + \psi^2 \sigma_q^2)^{-1/2} - \psi^2 \sigma_q^3 (1 + \psi^2 \sigma_q^2)^{-3/2}, \\
\frac{\partial \beta_{r,j}}{\partial \psi} &= \frac{\partial \sigma_r}{\partial \gamma_j} = \frac{\partial \rho}{\partial \gamma_j} = \frac{\partial \rho}{\partial \lambda} = 0. & \text{for } j = 1, \dots, p_r
\end{aligned}$$

Equations (3.4.1) and (3.4.2) then provide a measure of precision for the referral price parameter estimates, which corrects for the extra variation introduced by two stage estimation, whereby the quote price residuals are used in place of the true noise terms.

### 3.4.2 Standard errors of predictions

Standard errors for predictions of willingness to pay are also found by the Delta Method. For a quote with covariates  $\mathbf{x}_r$ , quote price  $q$ , and estimated residual  $\hat{\varepsilon}_q$ , willingness to pay is estimated by:

$$\hat{\pi}(q; \mathbf{x}) = \Phi \left( \mathbf{x}_r' \hat{\boldsymbol{\gamma}} + \hat{\lambda} q + \hat{\psi} \hat{\varepsilon}_q \right). \quad (3.4.3)$$

The variance of this prediction is given by:

$$\text{Var}(\hat{\pi}(q; \mathbf{x})) = \mathbf{w}(\hat{\phi}_r)' \Sigma_{\phi_r}^* \mathbf{w}(\hat{\phi}_r). \quad (3.4.4)$$

$\Sigma_{\phi_r}^*$  is the variance of the probit coefficients, corrected in the manner described above, and  $\mathbf{w}(\hat{\phi}_r)$  is a vector of partial derivatives of  $\hat{\pi}(q; \mathbf{x})$  w.r.t.  $\phi_r$ , evaluated at the MLEs:

$$\mathbf{w}(\phi_r) = \begin{pmatrix} \frac{\partial \hat{\pi}}{\partial \gamma} \\ \frac{\partial \hat{\pi}}{\partial \lambda} \\ \frac{\partial \hat{\pi}}{\partial \psi} \end{pmatrix}$$

where,

$$\begin{aligned} \frac{\partial \hat{\pi}}{\partial \gamma_j} &= x_j \phi(\mathbf{x}_r' \boldsymbol{\gamma} + \lambda q + \psi \hat{\varepsilon}_q), & \text{for } j = 1, \dots, p_r \\ \frac{\partial \hat{\pi}}{\partial \lambda} &= q \phi(\mathbf{x}_r' \boldsymbol{\gamma} + \lambda q + \psi \hat{\varepsilon}_q), \\ \frac{\partial \hat{\pi}}{\partial \psi} &= \hat{\varepsilon}_q \phi(\mathbf{x}_r' \boldsymbol{\gamma} + \lambda q + \psi \hat{\varepsilon}_q). \end{aligned}$$

When predictions are made on the probit-scale, both sides of (3.4.3) are transformed by  $\Phi^{-1}(\cdot)$ . The variance of the prediction is estimated in much the same way, except that the  $\phi(\cdot)$  terms in  $\mathbf{w}(\phi_r)$  are replaced by one.

### 3.5 Simulated example

A simple example with simulated data is used to demonstrate endogeneity and instrumental variable estimation for the model. Quote price and referral price are simulated

as follows:

$$\begin{aligned} q &= \beta_{q,0} + \beta_{q,1}x_c + \beta_{q,2}x_q + \varepsilon_q; & \varepsilon_q &\sim N(0, \sigma_q^2) \\ r &= \beta_{r,0} + \beta_{r,1}x_c + \varepsilon_r; & \varepsilon_r &\sim N(0, \sigma_r^2) \end{aligned}$$

A representative sample of 1,000 observations was generated, although typically much larger sample sizes are available in observational studies.

Both the joint covariate  $x_c$  and instrumental variable,  $x_q$ , were sampled independently from a Uniform(0,1) and parameter values used to generate the data are given in Tables 3.5.1 and 3.5.2. Figure 3.5.1 shows that the joint dependence of quote price and referral price on covariate  $x_c$  gives the appearance of an upward-sloping willingness to pay (WTP) curve and positive price elasticity. Both  $q$  and  $r$  are increasing functions of  $x_c$  with some added noise; however for low values of  $x_c$  sales agents generally overestimate the effect of  $x_c$  on referral price, whilst for high values of  $x_c$  agents underestimate the relationship. The result is that for low quotes, quote price tends to exceed referral price and these quotes are mostly rejected, whereas for high quotes, referral price tends to exceed quote price and these quotes are mostly accepted. In the plot on the far right, quotes have been divided into bins according to price (based on deciles) and the average price for each bin is plotted against the proportion of quotes which were accepted. The curve slopes upwards, giving the appearance of positive price elasticity.

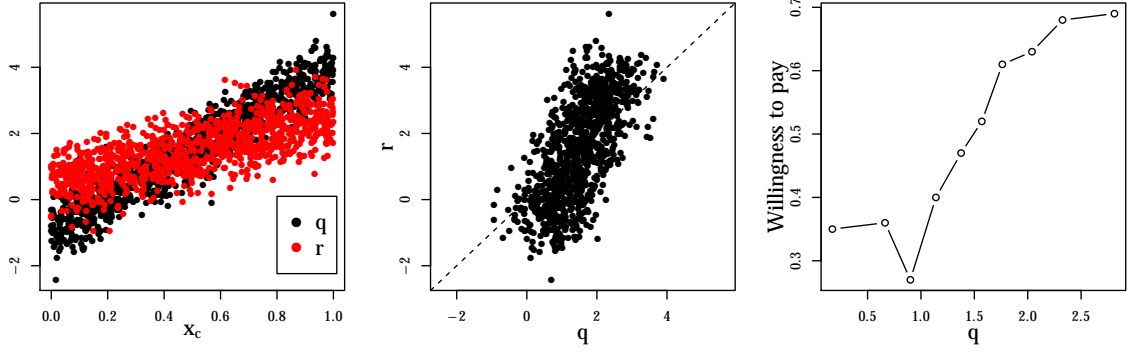


Figure 3.5.1: Simulated example of quote and referral price as a function of unobserved covariate  $x_c$ , resulting in an upward-sloping WTP curve (right). A (0,1) line is included in the centre plot for reference (.....).

Three models were fitted to the simulated data:

(1) Basic model

$$q = \hat{\beta}_{q,0} + \hat{\varepsilon}_q; \quad \hat{\varepsilon}_q \sim N(0, \hat{\sigma}_q^2).$$

$$\Phi(\hat{\pi}(q; \mathbf{x})) = \frac{\hat{\beta}_{r,0}}{\hat{\sigma}_r} - \frac{1}{\hat{\sigma}_r} q.$$

(2) IV model

$$q = \hat{\beta}_{q,0} + \hat{\beta}_{q,2} x_q + \hat{\varepsilon}_q; \quad \hat{\varepsilon}_q \sim N(0, \hat{\sigma}_q^2).$$

$$\Phi(\hat{\pi}(q; \mathbf{x})) = \frac{1}{\sqrt{1 - \hat{\rho}^2}} \left( \frac{\hat{\beta}_{r,0}}{\hat{\sigma}_r} - \frac{1}{\hat{\sigma}_r} q + \frac{\hat{\rho}}{\hat{\sigma}_q} \hat{\varepsilon}_q \right).$$

Table 3.5.1: Estimates of mean parameters for simulated example

	$\beta_{q,0}$	$\beta_{q,1}$	$\beta_{q,2}$	$\beta_{r,0}$	$\beta_{r,1}$
Basic model	1.46 (1.43, 1.52)	-	-	1.49 (1.37, 1.61)	-
IV model	1.03 (0.94, 1.12)	-	0.75 (0.75, 1.08)	1.34 (1.34, 1.60)	-
Full model	0.02 (-0.06, 0.10)	2.00 (1.90, 2.11)	0.96 (0.85, 1.07)	-1.16 (-1.69, -0.63)	5.32 (4.28, 6.37)
True	0	2	1	-1	5

(3) Full model

$$q = \hat{\beta}_{q,0} + \hat{\beta}_{q,1}x_c + \hat{\beta}_{q,2}x_q + \hat{\varepsilon}_q; \quad \hat{\varepsilon}_q \sim N(0, \hat{\sigma}_q^2).$$

$$\Phi(\hat{\pi}(q; \mathbf{x})) = \frac{1}{\sqrt{1 - \hat{\rho}^2}} \left( \frac{\hat{\beta}_{r,0}}{\hat{\sigma}_r} + \frac{\hat{\beta}_{r,1}}{\hat{\sigma}_r}x_c - \frac{1}{\hat{\sigma}_r}q + \frac{\hat{\rho}}{\hat{\sigma}_q}\hat{\varepsilon}_q \right).$$

In the first two models, the joint covariate  $x_c$  is treated as unobserved. The basic model therefore excludes  $x_c$  from both the quote price and referral price models, and also excludes the instrumental variable  $x_q$ . This provides a naive estimate of the WTP function, since quote price is assumed to be independent of willingness to pay. In the instrumental variables model,  $x_q$  is introduced to the quote price model. The residuals from this model are then entered into the probit regression to correct for endogeneity between quote price and referral price caused by  $x_c$ . The full model is a true representation of the data generating process and includes both covariates.

Tables 3.5.1 and 3.5.2 provide the parameter estimates from each model, alongside 95% confidence intervals and the true parameter values. In the basic model, the quote price residuals are not included in the WTP function. Therefore the procedure outlined in Section 3.4.1 for correcting standard errors of the referral price parameters

Table 3.5.2: Estimates of variance parameters for simulated example

	$\sigma_q$	$\sigma_r$	$\rho$
Basic model	0.80 (0.73, 0.88)	-2.55 (-2.67, -2.52)	- -
IV model	0.76 (0.70, 0.83)	1.51 (1.03, 1.98)	0.78 (0.68, 0.89)
Full model	0.49 (0.45, 0.54)	0.51 (0.32, 0.69)	0.01 (-0.25, 0.24)
True	0.5	0.5	0

is unnecessary. The standard errors of the parameters are taken directly from the inverse of the Observed Fisher Information, and confidence intervals are based on a normal approximation to the distribution of the estimators. However for the IV and full model, a correction must be applied to the standard errors of  $\{\hat{\beta}_{r,0}, \hat{\beta}_{r,1}, \hat{\sigma}_r, \hat{\rho}\}$  for using  $\hat{\varepsilon}_q$  in place of the true noise terms. The bootstrap method was used, in which observations are repeatedly sampled with replacement and the model is refitted. In this case the data was resampled 1,000 times. The variance of the bootstrap parameter estimates is then added to the estimated variance from the original model fit. The sum of these terms gives the corrected standard errors, which were then used to construct confidence intervals in the standard way.

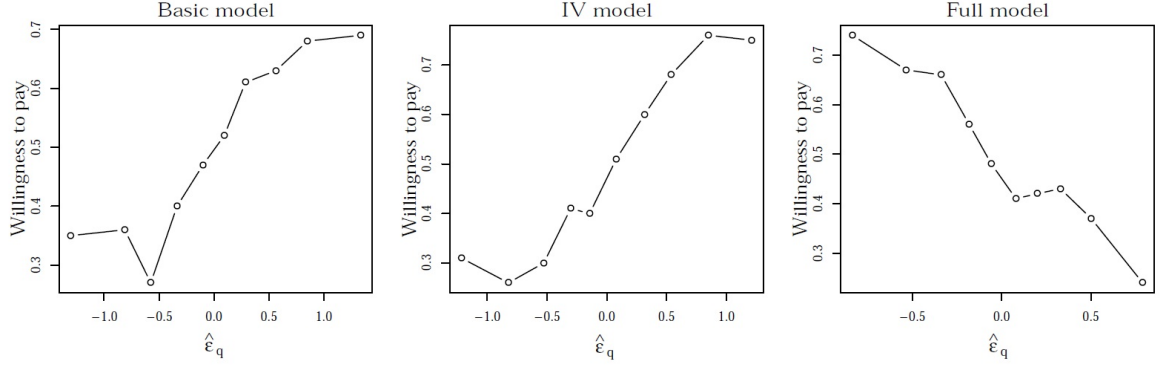


Figure 3.5.2: The relationship between willingness to pay and the estimated quote price residuals from each model.

Figure 3.5.2 shows the relationship between willingness to pay and the estimated price residuals from each model. In the basic model,  $\hat{\varepsilon}_q$  is effectively quote price and the relationship is identical to that shown in Figure 3.5.1. Since the common covariate  $x_c$  is missing from the model, the parameter estimates suffer from omitted variable bias. The coefficient for price in WTP function is positive and results in an invalid estimate for  $\sigma_r$  of -2.55.

The instrumental variable model includes  $x_q$ , so that the quote price residuals contain the true noise terms,  $\varepsilon_q$ , and the effect of  $x_c$ . Then since  $x_q$  is correlated with  $q$ , but independent of  $\hat{\varepsilon}_q$  and  $r$ , the inclusion of both  $q$  and  $\hat{\varepsilon}_q$  in the WTP function allows us to identify changes in willingness to pay produced by  $q$ , which are independent of  $x_c$ . The co-dependence of quote price and referral price on  $x_c$  is captured by residual correlation and we obtain a more sensible estimate for  $\sigma_r$ . The centre plot of Figure 3.5.1 shows that the proportion of acceptances increases with  $\hat{\varepsilon}_q$ , and the parameter estimates reflect this positive price elasticity, as  $\hat{\rho} > \hat{\sigma}_q/\hat{\sigma}_r$ .

The full model captures all variation in quote price and referral price produced

by the covariates. The residuals from the quote price model represent variation in  $q$  which is independent of  $x_c$  and hence also independent of referral price. A negative relationship therefore emerges between the proportion of accepted quotes and  $\hat{\varepsilon}_q$ , as a large noise term indicates a higher than average quote price (for a given level of  $x_c$ ) and is more likely to be rejected. Confidence intervals for the parameters contain the true values and  $\hat{\rho} \approx 0$ .

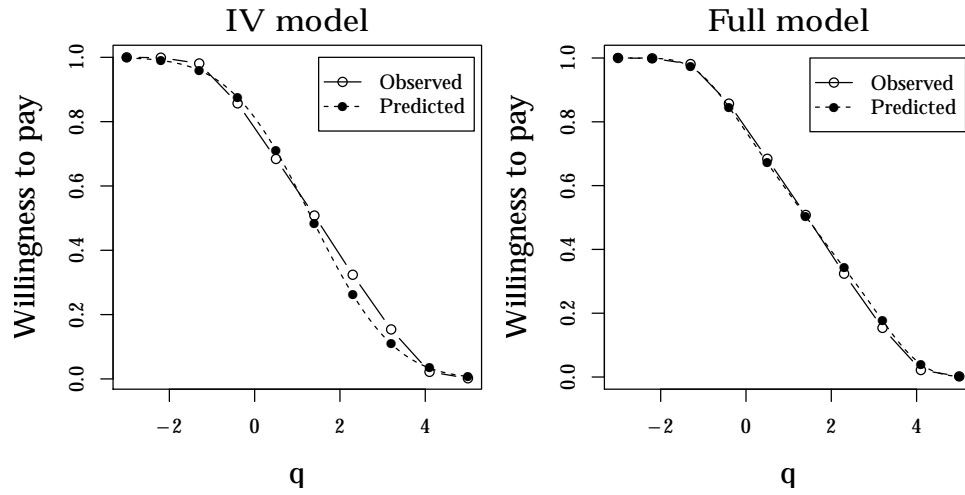


Figure 3.5.3: Predictions of willingness to pay for a new sample, alongside observed acceptance rates for the IV and full model.

Figure 3.5.3 shows acceptance rates for a new sample of quotes, generated in the same way as the training sample. Observed willingness to pay gives the proportion of quotes in which referral price exceeds a given value for  $q$ . Predictions of willingness to pay were made from the estimates for the IV model and full model. Since price is now fixed and no longer a random variable, willingness to pay is estimated from (3.1.5) as follows:



IV model

$$\Phi(\hat{\pi}(q; \mathbf{x})) = \frac{\hat{\beta}_{r,0}}{\hat{\sigma}_r} - \frac{1}{\hat{\sigma}_r} q.$$

Full model

$$\Phi(\hat{\pi}(q; \mathbf{x})) = \frac{\hat{\beta}_{r,0}}{\hat{\sigma}_r} + \frac{\hat{\beta}_{r,1}}{\hat{\sigma}_r} x_c - \frac{1}{\hat{\sigma}_r} q.$$

Under the full model  $x_c$  is assumed to have been observed, and so for a given value of  $q$ , willingness pay was predicted for all values of  $x_c$  observed in the new sample and then averaged. Estimates of willingness to pay are close to the observed values, as expected. Although  $x_c$  is omitted from the IV model, the higher estimate for  $\sigma_r$  accounts for this extra source of variation in referral price. Hence the effect of price on willingness to pay is accurately captured by the model.

## 3.6 Modelling considerations

In this section we introduce some possible extensions to the willingness to pay model, including a transformation for price, random effects, and non-linear covariate effects.

### 3.6.1 Treatment for non-normality

Since prices are always positive, the normality assumption on  $\varepsilon_q$  may not be sufficiently close to the truth and this may invalidate inferences drawn from the fitted model. In most cases and particularly in our application, variation in quote prices is small relative to the average price and hence the constraint for price to be positive is not violated by assumptions of normality. However, in instances where a normal dis-

tribution for  $\varepsilon_q$  is not supported by the data, for example due to skewness, price can be transformed into a variable for which a normal assumption is more appropriate.

One such example is the Box-Cox transformation (Box and Cox, 1964), which can be used to define a new, transformed price:

$$q^* = f(q) = \frac{q^\eta - 1}{\eta}. \quad (3.6.1)$$

Provided that  $f$  defines a monotonically increasing transformation, the procedure described in Section 3.3.2 remains unchanged, as the same transformation can be applied to referral price,  $r^* = f(r)$ . There are no theoretical implications for the model, since  $q^*$  and  $r^*$  can be assumed to have a joint normal distribution and the willingness to pay function is defined by  $P(R^* > q^* | q^*)$  rather than  $P(R > q | q)$ . Correction for endogeneity proceeds in the same way, and the probit regression is performed on  $q^*$  and the associated residual,  $\hat{\varepsilon}_q^*$ .

The transformation should have no tangible consequences for goodness of fit of the probit model, (3.3.10), since referral price is a latent variable. In order to confirm this, a Hosmer-Lemeshow test is performed on the standard model, as originally presented in Section 3.3.2, and also on the optimally transformed model, which uses  $q^*$  and  $r^*$ . Details of the Hosmer-Lemeshow test are provided in the next section. Given this assumption,  $\eta$  is chosen to maximise the pseudo-marginal likelihood for quote price:

$$L(\boldsymbol{\theta}_q, \eta | \mathbf{q}, \mathbf{X}) = \prod_{i=1}^n \frac{1}{\sigma_q} \phi \left( \frac{q_i^* - \mathbf{x}_{q,i}' \boldsymbol{\beta}_q}{\sigma_q} \right) q_i^{(\eta-1)},$$

where the last term comes from the Jacobian of the transformation. It is straightfor-

ward to compile a profile likelihood for  $\eta$ ; first by fixing  $\eta$  to some value and fitting the quote price model to the transformed price variable, then calculating the above likelihood value and repeating this for various values of  $\eta$ .

### 3.6.2 Measuring goodness of fit

Goodness of fit of both the quote price model and referral price model can be measured by Akaike's Information Criterion (AIC) or Bayesian Information Criterion (BIC). Both measures are based on the likelihood function and penalise model complexity:

$$\text{AIC} = -2\ell + 2p, \quad (3.6.2)$$

$$\text{BIC} = -2\ell + p \log(n), \quad (3.6.3)$$

where  $\ell$  is the log-likelihood value at the MLEs,  $p$  is the number of parameters, and  $n$  is the sample size. A model with a lower AIC or BIC value provides a better fit to the data than a model with higher values; and a larger difference in either AIC or BIC indicates stronger evidence for one model over the other. BIC values parsimony more strongly than AIC as it has a larger penalty for the number of model parameters, whereas AIC has a non-zero probability of overfitting, even for large samples (Hughes and King, 2003). For this reason, BIC is preferred for model selection here.

A Bayes factor measures the strength of evidence that a dataset was generated by one model over another, by the ratio of the marginal likelihoods of each model (see Section 2.4 for a definition of marginal likelihood). Model selection based on BIC uses the result that when sample size is large, the difference in BIC values between

two models can be used to approximate the Bayes factor (Kass and Raftery, 1995). Then in selecting between two models with BIC values of  $BIC_1$  and  $BIC_2$ , where  $BIC_1 < BIC_2$ , if  $BIC_2 - BIC_1 > 6$  there is strong evidence against Model 2 as a suitable model for the data, compared to Model 1.

The Hosmer-Lemeshow test is a method of assessing goodness of fit for models with a binary response variable (Hosmer and Lemeshow, 1980). Observations are ordered by their predicted probabilities and then partitioned into  $G$  groups of roughly equal size. For each group, the observed ( $O_g$ ) and expected ( $E_g$ ) number of successes is computed, where the latter is given by the sum of the fitted probabilities. These are then used to calculate the test statistic:

$$T = \sum_{g=1}^G \frac{(O_g - E_g)^2}{E_g \left(1 - \frac{E_g}{n_g}\right)},$$

where  $n_g$  denotes the number of observations in group  $g$ . Then provided that  $(p+1) < G$ , the test statistic approximately follows a Chi-square distribution with  $G-2$  degrees of freedom when the model is specified correctly. Hence a significant p-value ( $< 0.05$ ) is indicative of poor model fit.

### 3.6.3 Non-linear covariate effects

In some circumstances, a non-linear relationship between a covariate and response can be represented in a linear fashion by transforming the covariate. For example, log-transformation of positively skewed and large scale variables is a common practice. However, when the non-linear relationship cannot be represented by a simple

transformation, or by a polynomial of reasonable degree ( $<4$ ), a piecewise linear transformation can be used. Here, the domain of the covariate is divided into segments and a linear function is fitted between the covariate and response in each segment. The segments are defined by knots,  $\mathbf{x}_k = (x^{(1)}, x^{(2)}, \dots, x^{(s)})$ , and the transform is denoted by:

$$k(x, \mathbf{x}_k) = \begin{pmatrix} x \\ (x - x^{(1)}) \cdot \mathbb{I}[x > x^{(1)}] \\ \vdots \\ (x - x^{(s)}) \cdot \mathbb{I}[x > x^{(s)}] \end{pmatrix}. \quad (3.6.4)$$

The vector,  $k(x, \mathbf{x}_k)$ , is calculated for each observation and included in the model as a regressor. The corresponding set of coefficients,  $\boldsymbol{\beta} = (\beta^{(0)}, \beta^{(1)}, \dots, \beta^{(s)})$  defines the slope of the line in each segment, and is estimated in the standard way. For  $x \in [x^{(A)}, x^{(A+1)})$ , the slope of the line is given by  $\sum_{a=0}^A \beta^{(a)}$ . Selection of the number of knots,  $s$ , and knot locations,  $\mathbf{x}_k$ , is a matter of model selection. To avoid overfitting by introducing unnecessary knots, K-fold cross validation is used, whereby the sample is divided into  $K$  sub-samples of equal size. Generally  $K$  is chosen as 10. For a given number of knots, the knot locations are selected by fitting to the full sample and minimising BIC. Then the first sub-sample is removed from the sample and the model is refitted, after which the mean square error (MSE) is calculated for the removed sub-sample. This is repeated for all  $K$  sub-samples and the MSE is averaged. Then  $s$  is selected as the number of knots which produces the lowest average MSE.

The above linear function and other piecewise polynomial functions are jointly

referred to as regression splines, which allow for flexible specification of the relationship between a covariate and response. The most popular, a cubic spline, is continuous up to a second derivatives and so models the response as a smooth function of the predictor variable. The degree of smoothing is controlled by the basis dimension, which is effectively controlled by the number of knots specified for the spline. However for penalised regression splines, smoothness is controlled by introducing a penalty term to the objective function, which is proportional to the integrated square of the second derivative of the spline. Then provided that the number of knots is overspecified, neither the exact number of knots nor the knot locations have a great deal of influence on model fit (Wood, 2006, p.128). Smoothness is therefore determined by the size of the penalty term or the *smoothness parameter*,  $\tau$ , and as  $\tau \rightarrow \infty$ , the spline tends to a straight line. The smoothness parameter is estimated by minimising the Generalised Cross-Validation (GCV) score, which is equivalent to the expected squared error in predicting a new variable.

### 3.6.4 Random effects

In addition to the effects of covariates, random effects are a useful approach to modelling unobserved heterogeneity or excess variability between the units or subjects in a sample. They account for correlation between the responses for a particular subject or unit that persists even after removing the effects of covariates (Collett, 2002). The standard mixed effects linear model is described by:

$$\mathbf{q} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon},$$

where  $\varepsilon_1, \dots, \varepsilon_n$  are iid  $N(0, \sigma^2)$  and  $b_1, \dots, b_m$  are iid  $N(0, \sigma_b^2)$ . The regression parameters are made up of the population-specific parameters,  $\beta$ ,  $\sigma^2$  and  $\sigma_b^2$ , and the subject-specific effects  $\mathbf{b}$ . The  $(n \times m)$  matrix  $Z$  is comprised of 0's and 1's, where each row contains a single non-zero entry indicating the relevant subject for the observation. Covariates with a fixed effect on the response are included in the  $(n \times p)$  matrix  $\mathbf{X}$ .

The random effect captures systematic deviations from the expected value of the response, which are not accounted for by the fixed effects. For example, a customer random effect in the quote price model would mean that the prices offered to a particular customer tend to deviate from the average by a consistent amount. For customer  $j$ , this amount is  $b_j$ . Then the residuals  $\varepsilon_j$  are independent of each other and of all other sources of variation in the customer's quote prices. If a random effect is omitted and residual correlation exists between the responses of each subject, estimates of the fixed effects will be biased and inconsistent.

The parameters  $\sigma^2$  and  $\sigma_b^2$  are referred to as *variance components*. The fixed effects parameters,  $\beta$ , and variance components can be estimated by ML estimation, for which the random effects are integrated out of the model. This introduces some bias into estimates of the variance components, due to the loss of degrees of freedom involved with estimating the fixed effects (Verbeke and Molenberghs, 2000). Restricted Maximum Likelihood Estimation (REML) circumvents this bias by maximising the likelihood of some error contrasts,  $\mathbf{u}$ . Each  $\mathbf{u}$  is a linear combination of  $\mathbf{q}$  with expectation zero, and hence the distribution of  $\mathbf{u}$  is independent of the fixed effects. However, since a change in the fixed effects leads to a different set of error contrasts,

the likelihoods of two models with different specifications of fixed effects are not directly comparable (Faraway, 2004). For this reason, and also because the bias in estimates of variance components will be small for large samples, we use maximum likelihood estimation.

Estimates of the random effects themselves are often of interest, since they allow us to identify individuals whose behaviour deviates most from the population average. In a willingness to pay context, customers with extreme random effects values are those that are most price-sensitive or price-insensitive and therefore the customers who are most likely to respond to customised pricing strategies. For customers with few quotes, estimates use ‘borrowing strength’ across customers which means that random effects are shrunk towards a common mean. A natural predictor of the random effects,  $\hat{\mathbf{b}}$ , is obtained by minimising the mean squared prediction error,  $\mathbb{E}[(\hat{\mathbf{b}} - \mathbf{b})'(\hat{\mathbf{b}} - \mathbf{b})]$ , where the expectation is taken over both  $\mathbf{b}$  and  $\mathbf{q}$ . This gives the Best Linear Unbiased Predictor (BLUP):

$$\hat{b}_j = \mathbb{E}(b_j|\mathbf{q}) = \frac{\hat{\sigma}_c^2}{\hat{\sigma}^2 + n_j \hat{\sigma}_c^2} \sum_{i \in C_j} \left( q_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}} \right), \quad (3.6.5)$$

where  $C_j$  is the set of indices corresponding to subject  $j$ , and  $n_j$  is the number of indices in this set. This is also referred to as empirical Bayes prediction, as (3.6.5) gives the mean of the empirical posterior distribution of  $b_j$ . The term ‘empirical’ is used because the distribution is conditional on estimates of the model parameters. The standard error of the prediction is given by:



$$\text{Var}(b_j|\mathbf{q}) = \frac{\hat{\sigma}_b^2 \hat{\sigma}^2}{\hat{\sigma}^2 + n_j \hat{\sigma}_b^2}. \quad (3.6.6)$$

Then (3.6.5) and (3.6.6) can be used to predict the responses of subjects in the training sample. For new subjects, predictions are made from just the fixed effects, which represent the population average.

Assessing the significance of random effects requires testing hypotheses of the form  $H_0 : \sigma_b = 0$  vs.  $H_1 : \sigma_b > 0$ , which are on the boundary of the parameter space and hence change the asymptotic distribution of some classical test statistics (Verbeke and Molenberghs, 2000). Several authors have studied the number of degrees of freedom to include in the penalty terms of the AIC and BIC for mixed models (Müller et al., 2013; Greven and Kneib, 2010; Delattre et al., 2014). Here we use the approach of Delattre et al. (2014):

$$\text{BIC} = -2\ell + p \log(n) + 2 \log(m),$$

where  $m$  is the number of subjects in the sample,  $p$  is the number of fixed effects and  $\ell$  is the maximised log-likelihood of the marginal model, in which the random effects have been integrated out.

A customer random effect in the probit regression would represent the idea that some individuals are more prone to accept a price than others. A probit model with mixed effects is denoted by:

$$\Phi^{-1}(\mathbf{p}) = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{d},$$

where  $p_i (i = 1, \dots, n)$  is the probability of a success and  $d_1, \dots, d_m$  are iid  $N(0, \sigma_d^2)$ . Then on the  $n_j$  occasions that customer  $j$  requests a quote, their referral price deviates from the expected referral price under a given set of covariates by  $d_j$ . The above generalised linear mixed effects model can be fitted by ML estimation, in which Gaussian quadrature is used to marginalise over the random effects. Generalisations of the empirical Bayes predictors (3.6.5) and (3.6.6) are available for the random effects (Skrondal and Rabe-Hesketh, 2009).

We have assumed here that the two customer effects - a random effect on the quote price,  $b_i$ , and a random effect on the referral price,  $d_i$  - are independent, but this need not be the case. It is also possible for the random effects to be correlated,

$$\begin{pmatrix} b_j \\ d_j \end{pmatrix} \sim N \left( \mathbf{0}, \begin{pmatrix} \sigma_b^2 & \omega\sigma_b\sigma_d \\ \omega\sigma_b\sigma_d & \sigma_d^2 \end{pmatrix} \right).$$

In this case, two stage estimation cannot be used, as each random effect cannot be integrated out of the model separately. Maximum likelihood estimation of the full model must be performed, after marginalising over both sets of random effects. With multiple random effects in the model, this quickly becomes a computationally intensive and complex procedure and so for convenience, we assume independence between the random effects of the quote price model and referral price model. Further details are provided in Sections 4.2.3 and 4.4.2. If a random effect on the quote price and a

random effect on the referral price are wrongly assumed to be independent, this will result in a higher residual correlation between the two variables.

# Chapter 4

## Heating Oil Case Study

Here we present an overview of the heating oil market and introduce the data available, highlighting some of the key relationships between variables. This is followed by an examination of empirical willingness to pay, the relationship between price and the proportion of acceptances observed in the sample data. We then provide a description of the process through which we created an appropriate model for the data, including the selection of predictors, transformations of the data and the introduction of random effects. The chapter concludes with a comparison of the model estimates to the standard willingness to pay curve in which instrumental variables are not included to correct for endogeneity.

### 4.1 The market for heating oil

Approximately 15% of UK households are not connected to the mains gas grid, and of these, 1.6 million use kerosene to heat their homes. Although mostly located in rural

areas these households are geographically dispersed, covering all social grades and a broad range of income profiles, and therefore customer heterogeneity is a prominent feature of the market (OFT, 2011). Kerosene is typically stored in a tank which is kept externally and customers purchase kerosene from a supplier who delivers the oil directly to the tank.

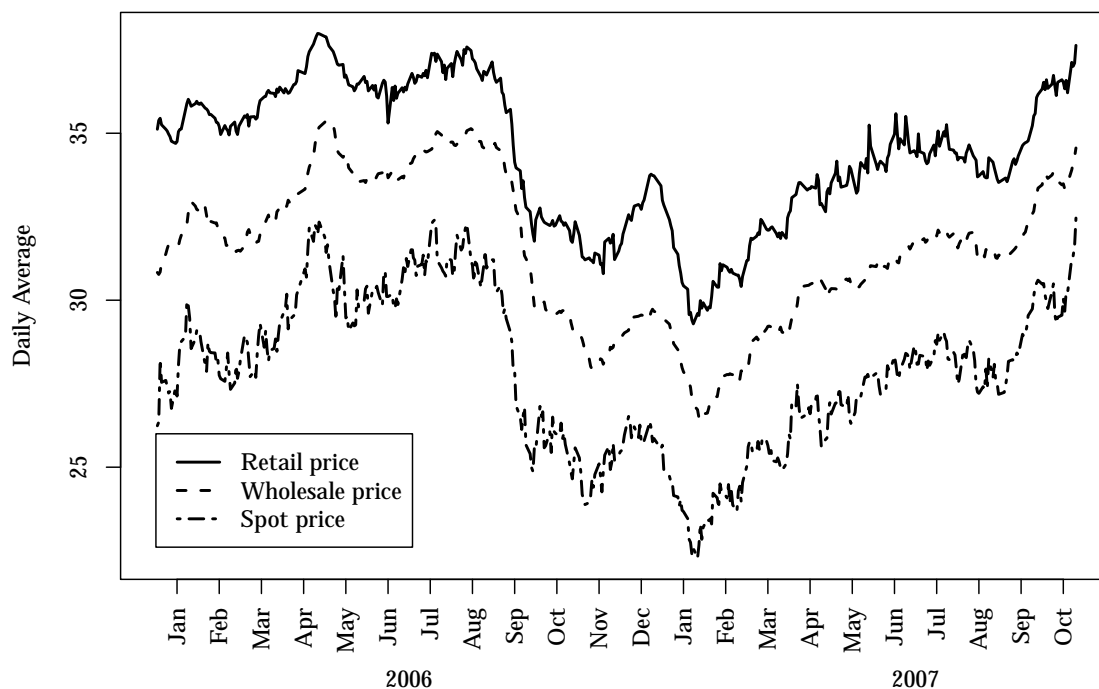


Figure 4.1.1: Daily average retail price, wholesale price and spot price of kerosene over the study period.

The retail price is the price quoted to the customer, in terms of pence per litre, and is comprised of two components: the wholesale price paid by the seller to purchase the oil, and the seller's mark-up. Figure 4.1.1 demonstrates that the retail price is heavily determined by the wholesale price, and the latter portion of the study period saw steadily rising inflation of prices. Kerosene is distilled from crude oil and around one quarter of UK kerosene production is used as heating oil, with the remainder being

used for jet fuel. As such, the wholesale price for heating oil is typically defined with reference to the international wholesale price of jet kerosene, referred to here as *spot price*; and spot price itself is determined by the price of crude oil (OFT, 2011, p.99). Volatility in crude oil prices means that the wholesale price can vary between  $\pm 10\%$  of the average in any given day. As a result, two identical customers who request quotes on separate days can expect to pay considerably different prices. Furthermore, the nature of the sales process, whereby prices are offered on a transaction-by-transaction basis, means that the retail price is also influenced by individual customer negotiations and short-term fluctuations in demand, as well as selective discounting.

#### 4.1.1 Description of dataset and data cleaning

The sample data contains the details of quotes made by a particular supplier over an 18 month period beginning in January 2006. Descriptions of the variables provided, as well as those created for modelling purposes, are shown in Table 4.1.1.

The seller operates 39 depots across the UK, and over the study period orders for deliveries from these depots were managed by a team of 302 sales agents. The vast majority of customers use the same depot for every order and a third of customers requested just one quote over the study period, whilst 23% requested two and 18% requested three. Most customers requested the same amount of fuel for each quote. For quotes with missing values for volume, these were estimated by the volume requested by the same customer in other quotes, provided that the orders were consistent. Some customers will ‘shop around’ and request multiple quotes within a short period of time. If a customer was found to have multiple quotes within a seven

Table 4.1.1: Variable definitions

Name	Description
y	binary variable indicating whether the quote was accepted or not
retailPrice*	price quoted to customer
wholePrice*	wholesale price of kerosene at the time of the quote
spotPrice*	international wholesale price of jet kerosene
margin*	retailPrice - spotPrice
cost*	wholePrice - spotPrice
cust	code representing customer requesting quote
agent	code representing sales agent responsible for the quote
depot	code representing depot from which the delivery was made
day	index representing date of quote
month	month in which the quote was requested
vol	volume of fuel requested (in L)
temp	average local temperature on the day prior to the quote (in °C)
postcode	first half of customer's postcode
dist	approximate distance (in km) between the customer and depot
quoteNum	no. of quotes requested by the customer from the beginning of the study period up to and including the present quote
quoteInt	no. of days between customer's present quote and their previous quote
quoteHist	code representing the customer's response to their last four quotes
depotCount	total no. of quotes requested at the depot on the day of the quote
volCount	total volume of fuel sold at the depot on the day of the quote (in L)
callTime	approximate duration of the call in minutes, calculated by inverting the no. of calls taken by the agent that day and multiplying by 480 (the no. of minutes in a working day)
shop	binary variable indicating whether a customer has been noted to request multiple quotes within a week (and therefore shop around)
quoted	binary variable indicating whether the agent offered a price which deviated from the reference price given by the CRM system

\*in pence per litre (ppl)

day period, they were labelled as shopping around. Since the quotes are related to the same order, only the last quote was kept in these instances, as the customer's final decision would have been determined on that basis.

Depot locations were approximated using customers' postcodes; the latitudes and longitudes associated with each postcode were averaged across all customers using a particular depot, and this formed a proxy for that depot's location. A number of recording errors were present for the variables `retailPrice` and `wholePrice`, for example where a decimal place had been misplaced. Given the magnitude of suspected outliers, and the tendency for these variables to remain relatively stable over the short-term whilst displaying a long-term trend, quotes were marked as being recording errors based on a moving average alongside a moving measure of spread. For each day of the study period  $d$ , an expected daily price  $\bar{x}_d$  and daily spread  $s_d^2$  were calculated from the set of quotes occurring within a 31 day interval centred on  $d$ ,  $D = \{\text{day} : (d - 15) \leq \text{day} \leq (d + 15)\}$ . This was given by the sample median and variance of  $x$  for the quotes in  $D$ , and a tolerance was set,  $\delta$ , based on visual inspection of the data. Then quotes occurring on day  $d$ , which fell outside of the range  $\bar{x}_d \pm \delta s_d$  were labelled as recording errors and removed.

Missing temperatures were interpolated from quotes requested on the same day, for the nearest customer. A handful of quotes were recorded on weekends and bank holidays and these were removed, since they occurred outside the normal operating hours of the business. Finally, quotes with missing information were removed and the remaining sample contained approximately 306,000 quotes for 111,000 customers.



### 4.1.2 Selection of price and cost variables

Before the willingness to pay model can be fitted, it is necessary to consider the process by which customers will form a referral price. Almost all customers are served by at least four suppliers, and although customer loyalty to a particular supplier is commonplace, customers are able to (and some indeed do) shop around (OFT, 2011).

In general, suppliers operating in the same area will be subject to the same wholesale price. For example, in pairwise comparisons between depots of the daily average wholesale price, the Pearson correlation coefficient ranged between 0.70 and 0.99. As a result it is the mark-up over the wholesale price, rather than the retail price itself, which encourages substitution between suppliers. However wholesale price is volatile and for any given depot, can vary by up to  $\pm 10\%$  around the daily average over the course of a day. A more realistic proposal is that customers' price expectations are formed relative to the spot price, which is widely published and fixed across the UK each day. It can therefore be argued that rather than mark-up over wholesale price, `retailPrice - wholePrice`, willingness to pay is determined by `retailPrice - spotPrice`.

The cost of the sale is typically an effective instrument for price, since it affects the price offered by the seller but not willingness to pay (Zhang et al., 2014; Fong et al., 2010). The plot on the left of Figure 4.1.2 shows mark-up over the wholesale price has a weak and slightly negative correlation with wholesale price. This suggests that changes to the seller's cost are not fully transferred to the customer. Likewise, the mark-up over the spot price is weakly correlated with spot price. In linear regressions

for both of these relationships, the  $R^2$  value was less than 0.01. On the other hand, if retail price and wholesale price are both standardised by subtracting spot price, a positive and stronger correlation emerges, with an  $R^2$  value of 0.06. An increase in the wholesale price over the spot price is passed on to the customer via a higher mark-up over the spot price. The latter relationship would therefore be more effective for endogeneity correction.

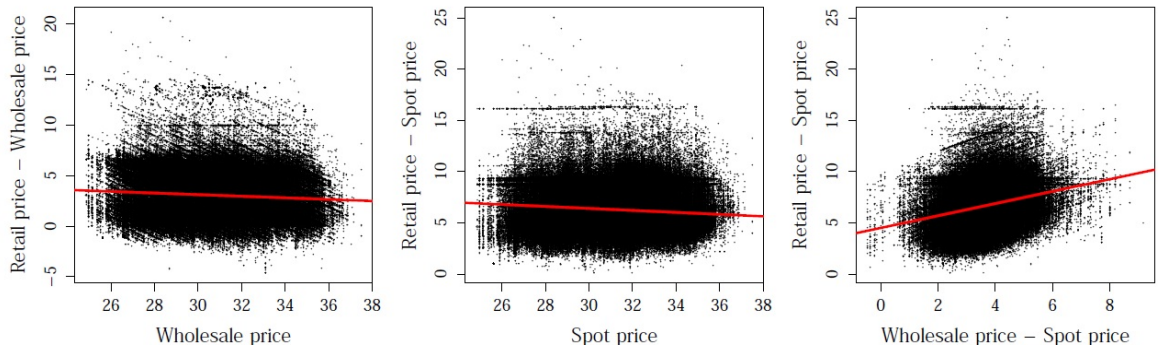


Figure 4.1.2: Different representations of the relationship between price and cost, based on retail price, wholesale price and spot price, with a line of best fit (—).

For these reasons, cost is defined by the difference between wholesale price and spot price; and our interest lies in quantifying the relationship between `margin` and willingness to pay, where margin is the mark-up over spot price. In using `margin` as the price variable, industry-wide cost levels are separated from deal-specific pricing decisions.

### 4.1.3 Exploratory data analysis

Seasonality of prices is demonstrated in Figure 4.1.3. Margins are generally lower during the summer months, when higher temperatures result in lower demand, and

they peak around January/December. There is also more variability in 2007 compared to the previous year.

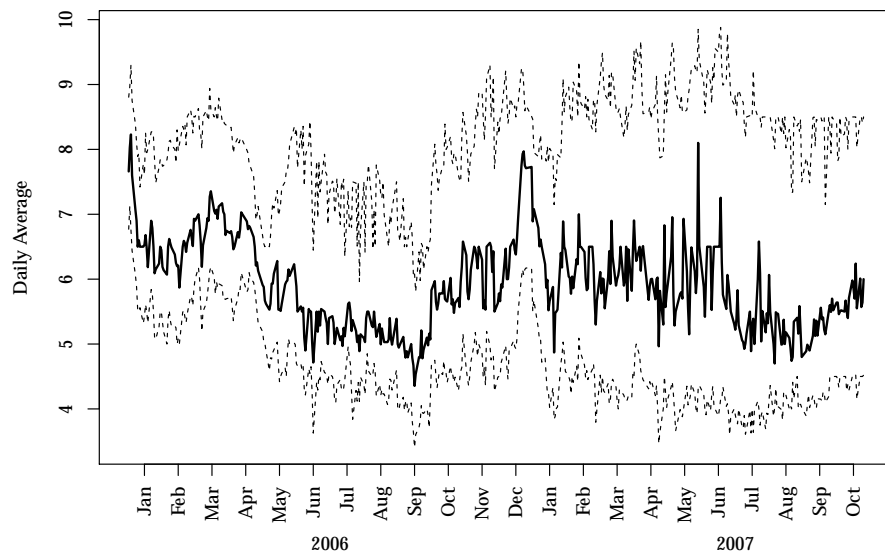


Figure 4.1.3: The median (—) and upper and lower quartiles (.....) of margins offered on each day of business.

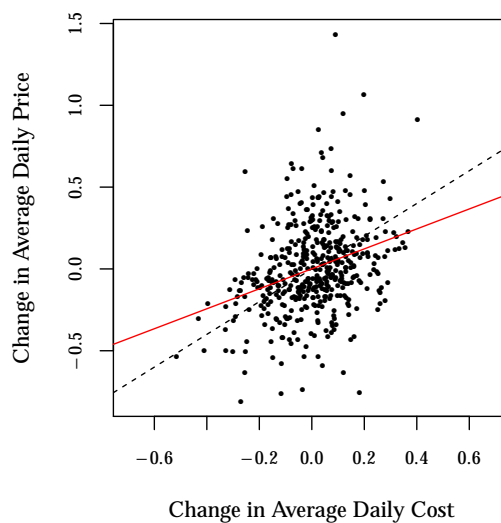


Figure 4.1.4: Changes in the average daily wholesale price and corresponding changes in average retail price, with a line of best fit (—) and (0, 1) line for reference (-----).

A widely held belief in econometrics is that prices do not move in the direction

or magnitude indicated by cost changes, referred to as *price asymmetry* (Elmaghraby et al., 2012). In particular, sellers are quicker to pass on cost increases to consumers than they are to respond to a fall in costs. There is no compelling evidence for this effect in Figure 4.1.4, which shows changes in the average daily wholesale price and retail price. Under price asymmetry, the slope of this graph is expected to exceed one, however price changes are generally more variable than cost changes and only a weak relationship emerges.

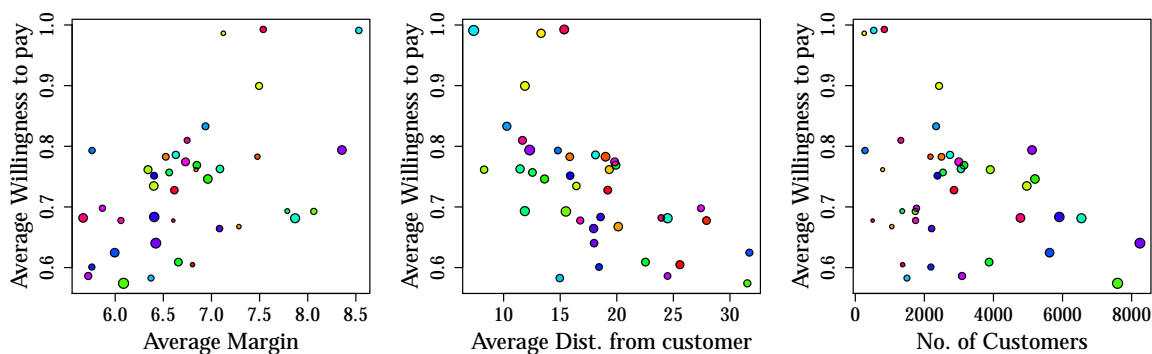


Figure 4.1.5: Average features for each depot. Points are colour coded by depot and in the left and right plots, point size is proportional to the number of quotes requested for each depot. In the centre plot, point size is proportional to average margin for each depot.

The size of depots varies considerably, with the largest having 8,000 customers and the smallest having 150 customers, and sales between these depots can differ by a factor of almost 60 (see Figure 4.1.6). As a result, pricing decisions and purchasing behaviour associated with each depot can vary from one depot to the next, as shown in Figure 4.1.5. Some of the smaller depots offer higher margins on average and yet also have higher acceptance rates. A possible explanation for this is that smaller depots have more loyal customers, since they are likely to be in rural locations with fewer

competitors. At the other extreme, large depots face stronger competitive pressure and offer lower margins to maintain a 60-70% acceptance rate.

Willingness to pay appears to be higher for depots with local customers, where distance from customer is small, despite these depots offering relatively higher margins. Whereas depots that are fairly isolated from their customers offer lower margins and also have lower willingness to pay. This suggests that customers living closer to a depot are less price sensitive. Figure 4.1.6 shows daily volume of sales for two depots. There is strong seasonality in sales at the larger depot, while sales at the smaller depot are relatively constant over time with long periods of zero sales, presumably because the depot was not operating.

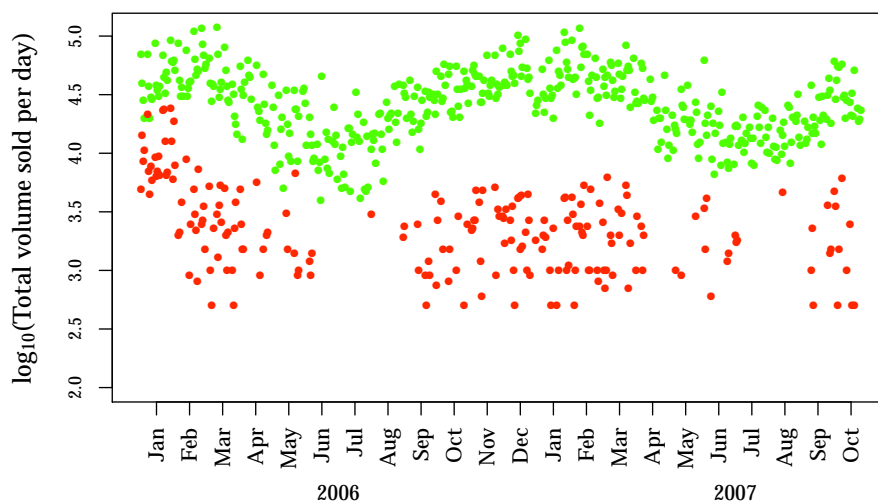


Figure 4.1.6: Total volume of fuel sold per day at one of the largest (—) and smallest (—) depots.

Spatial variation in prices can be expected as a result of regional differences in temperature and transportation costs. Figure 4.1.8 presents the average margin for each postcode on the busiest day in the study period, and indicates that prices tend

to be lower in North-East England, by almost 10 ppl.

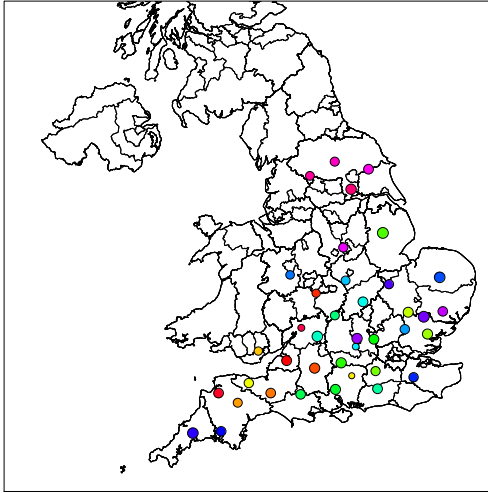


Figure 4.1.7: Approximate depot locations, where point size is proportional to the no. of customers for each depot.

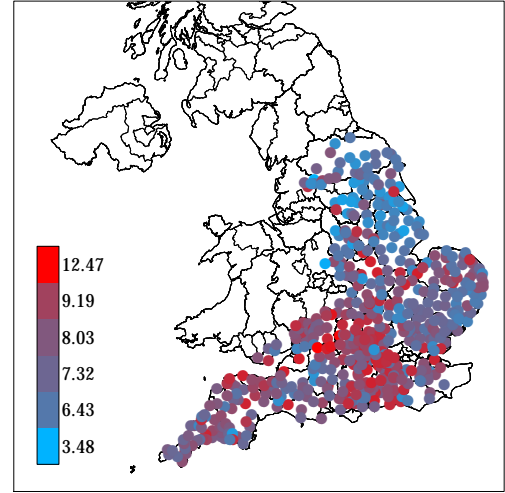


Figure 4.1.8: Average margin for each postcode on 27<sup>th</sup> Feb 2006, the busiest day of the study period.

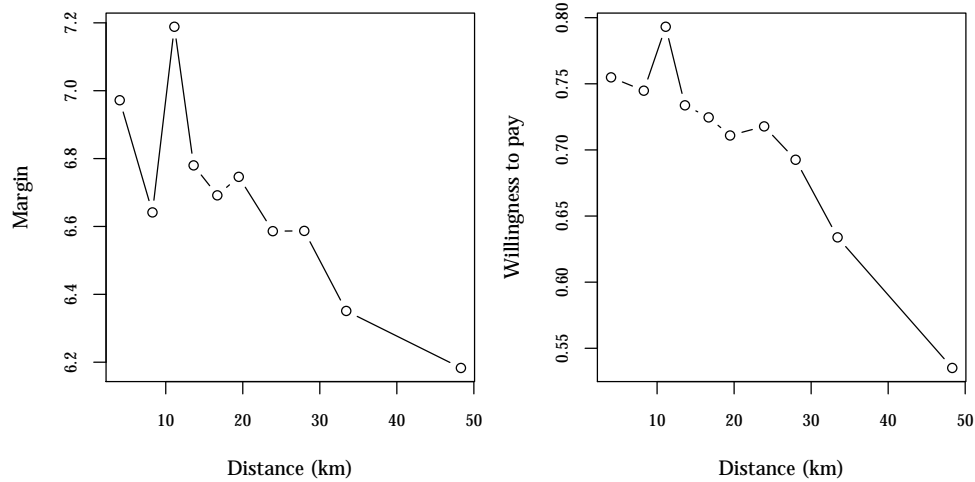


Figure 4.1.9: Relationship between `dist` and `margin` (left) and `dist` and `acc` (right). Observations have been binned according to quantiles of `dist`.

Used as a proxy for delivery and transportation costs, `dist` is the straight-line distance between the customer's postcode and the approximate location of the depot.

Figure 4.1.9 shows that variation in margin over the customer's distance from the

depot is small; however customers living further away are generally less likely to accept a quote, presumably because these customers will compare quotes with more local suppliers.

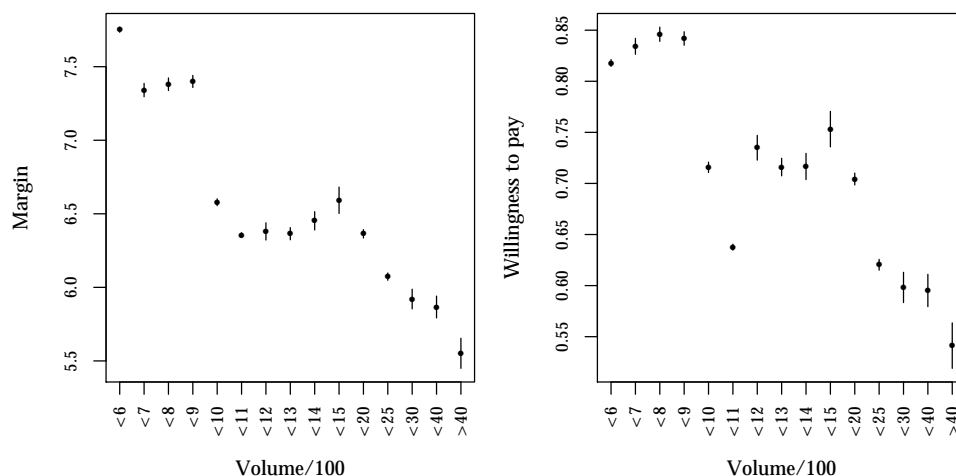


Figure 4.1.10: Average margin and acceptance rates (with 95% confidence intervals) across orders of different sizes. Quotes were divided into bins according to `volume`.

A key variable is the volume of fuel requested since many suppliers offer a quantity discount, and this effect was prominent in the sample data (Figure 4.1.10). A difference of over 2 ppl was found between the average margin offered for 500 litres (the minimum order size) compared to orders twice as large. A plot of volume against the proportion of sales also indicates that customers who request more fuel are more likely to reject their quotes; presumably because minor differences in the pence per litre quote will have a larger impact on the price of their total order and so the benefits of shopping around are greater. Figure 4.1.11 shows that three times as much fuel is sold on colder days compared to warmer days, which is consistent with the seasonality of price. However, orders on warmer days are larger by an average of 100 litres. Since some customers are known to use kerosene to heat swimming pools, we

can expect these customers to request more fuel over the summer. There may also be some strategic behaviour on the part of customers, who take advantage of lower prices by ordering more fuel.

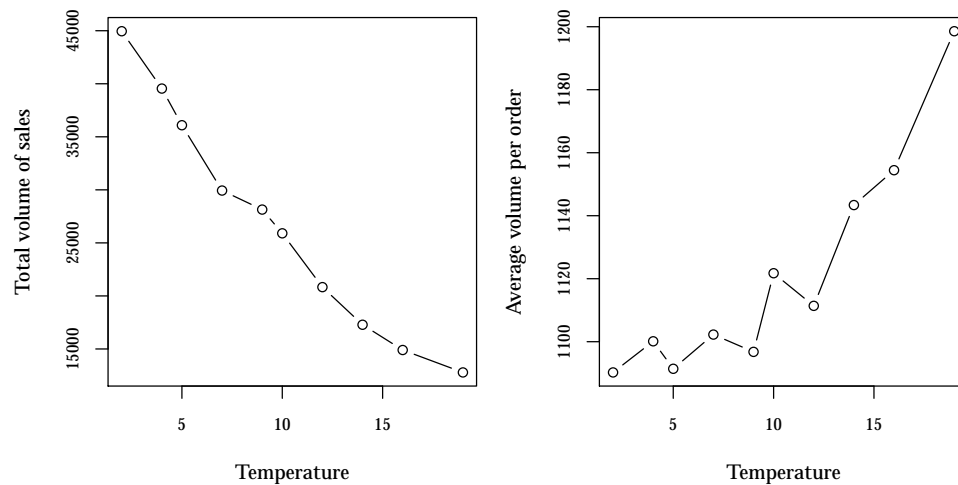


Figure 4.1.11: Relationship between volumes requested by customers and temperature. Quotes were divided into bins according to temperature.

Given that a sales agent has access to the customer's purchase history at the time of the quote, this is likely to have a strong influence on the price offered by the agent. Figure 4.1.12 shows how margin and willingness to pay varies according to the number of quotes the customer has previously requested in the sample. New customers (a quote number of one) are offered the lowest prices and are also least likely to accept the quote. Long-standing customers are charged higher margins on average, but despite this, they are also more likely to accept a quote.



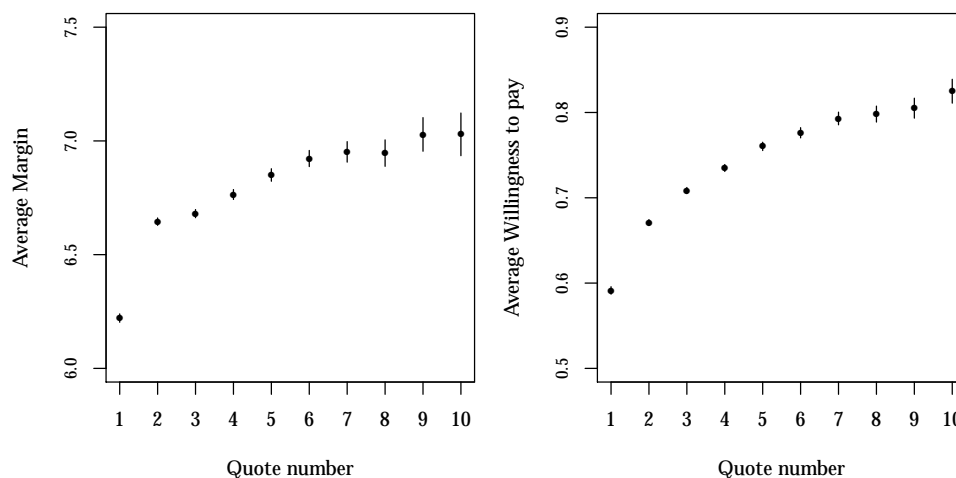


Figure 4.1.12: Average margin and acceptance rates (with 95% confidence intervals), according to the number of quotes the customer has previously requested.

However, if we examine purchase history at a more detailed level, it becomes apparent that Figure 4.1.12 does not capture the full extent of customer differences in price and willingness to pay. A customer's previous responses are represented by `quoteHist`, in which customers are divided into 31 categories depending on their responses to their four previous quotes. A quote may have been accepted (1) or rejected (0) or not occurred (E), hence customers in the 110E category have three previous quotes and accepted their last two quotes and rejected their first quote.

Figure 4.1.13 shows the average margin and willingness to pay for quotes that fall into each of these categories. Customers who call for a second time are offered margins 0.5p lower than new customers if they rejected their last quote, whereas customers that accepted are charged 0.8p more. A similar trend is found amongst customers with four or more quotes. Customers who accept twice in their four previous quotes are charged moderately higher margins than customers who accepted once, and moderately lower

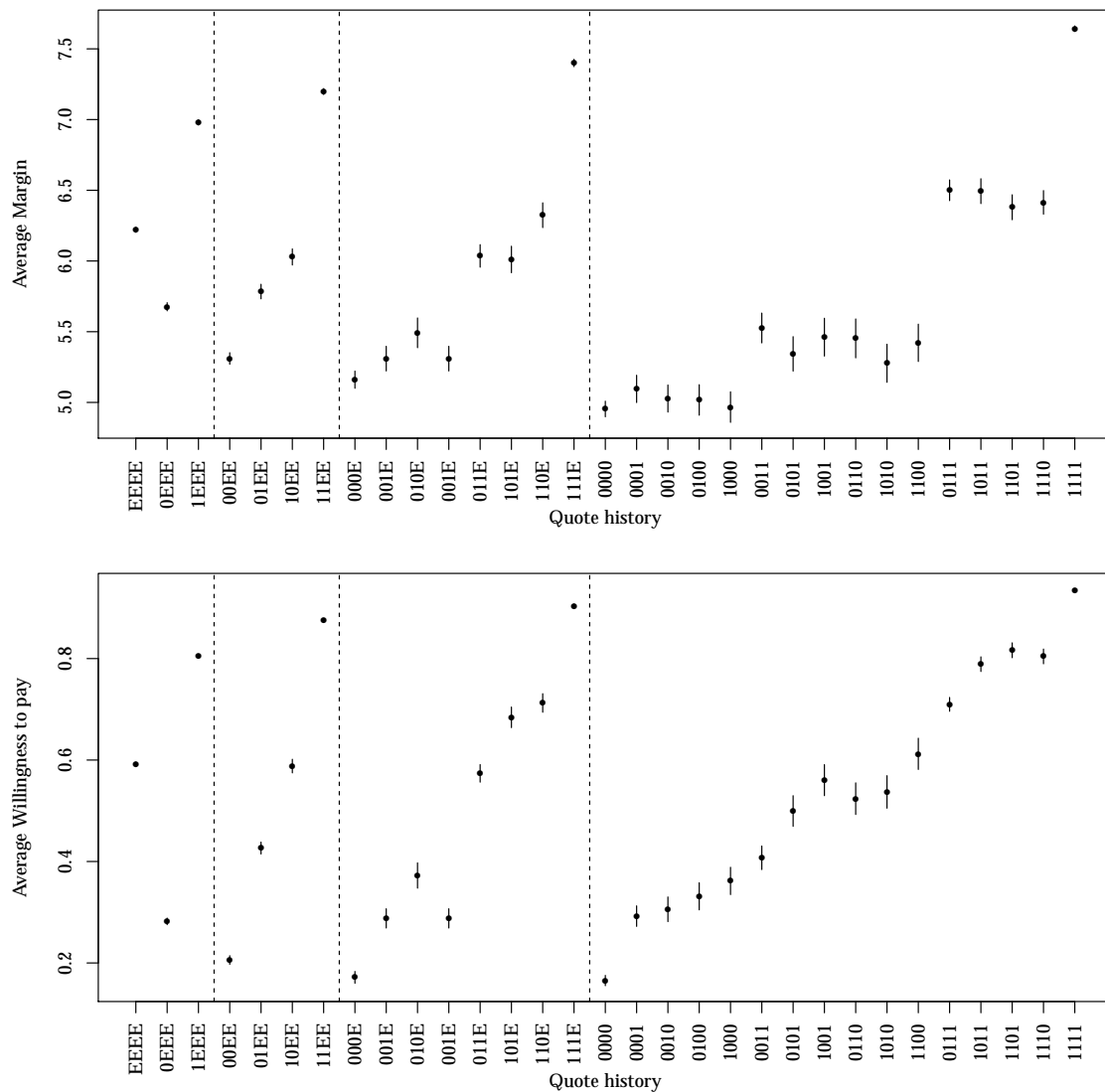


Figure 4.1.13: Average margin and acceptance rates (with 95% confidence intervals) across orders at different stages of the customer's purchase history. Categories have been grouped for easier interpretation: one or no previous quotes, two previous quotes, three previous quotes and four or more previous quotes.

margins than customers who accepted three times. Loyalty is penalised, rather than rewarded, as customers who always accept are offered the highest margins on average. A similar pattern is evident amongst the average acceptance rates for each category. It is also important to note that acceptance rates differ by how recently a customer's previous acceptance occurred. For example, all customers in categories 0001, 0010, 0100 and 1000 accepted one of their four previous quotes. However customers who accepted their most recent quote (1000) are more likely to accept their next quote, compared to customers in the other categories. The same affect does not appear to be present for margin.

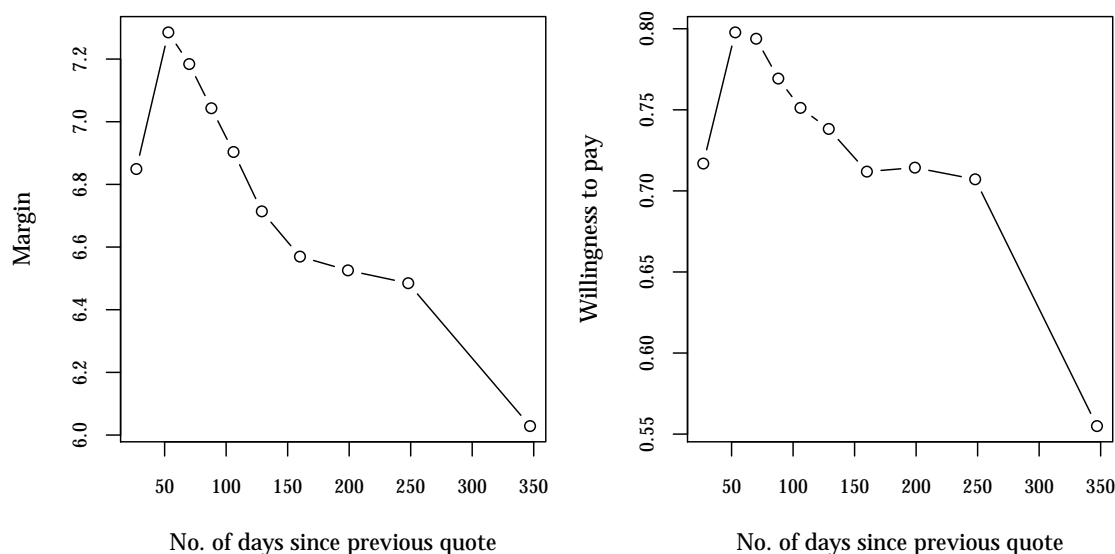


Figure 4.1.14: Relationship between `quoteInt` and `margin` (left) and `quoteInt` and `acc` (right). Observations have been binned according to the number of days since the customer last requested a quote.

More active customers can be defined as those who frequently contact the seller to request a quote. This is captured by the variable `quoteInt`, which counts the number of days since the customer previously requested a quote. Figure 4.1.14 shows

that customers who rarely request a quote are much less likely to accept, presumably because they tend to purchase fuel from other suppliers and are therefore price sensitive. Generally acceptance rates decrease as customers become less active, however customers who request a quote less than 50 days after their previous quote are not in keeping with this trend. This may be due to customers shopping around or delaying purchases in the hopes of lower future prices.

#### 4.1.4 Empirical willingness to pay

From the sample data we can examine empirical willingness to pay, or the relationship between margin and the proportion of sales. The curve indicates the presence of positive elasticity for intermediate margins, where the proportion of sales increases with margin. At a superficial level, this implies that customers who are offered 10 ppl are more likely to accept their quote than customers who are quoted 4 ppl. With few exceptions, products and services are generally subject to negative price elasticity, since the proportion of customers willing to pay a given price is a strictly decreasing function of price. Under the basic referral price framework introduced in Section 2.1.2, positive price elasticity implies a decreasing distribution function for referral prices, so that  $F_R(q) > F_R(q + h)$  for  $h > 0$ . Clearly then, price cannot be assumed to be independent of willingness to pay in this case.

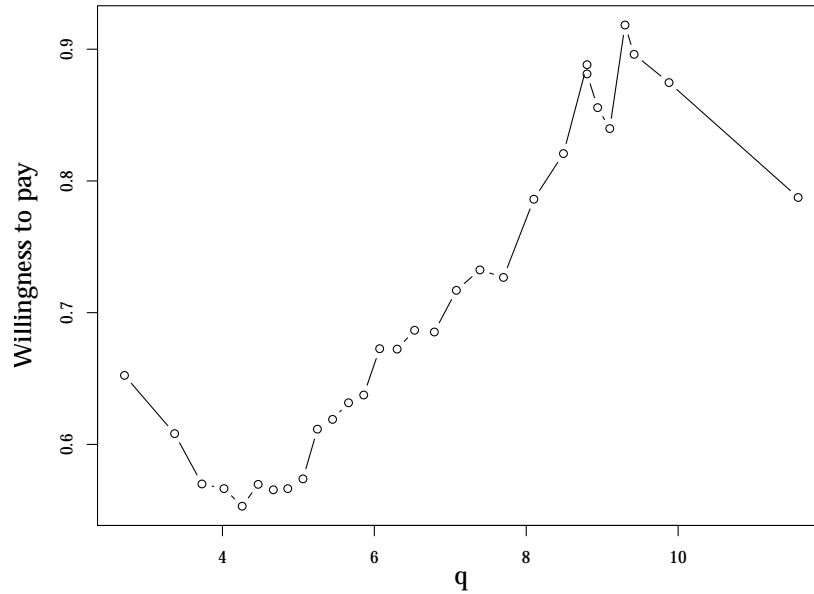


Figure 4.1.15: Proportion of quotes accepted for different intervals of margin, or  $q$ .

In Section 3.2 a simple example of positive price elasticity was demonstrated with simulated data. The same underlying relationships are likely to be responsible for the trend shown in Figure 4.1.15. Since quotes are offered in real-time to individual customers, the quote price  $q$  is essentially the sales agent's estimate of the customer's referral price,  $r$ . We would therefore expect some form of dependence between  $q$  and  $r$ , whereby agents are able to detect to some extent when a customer is willing to pay more; and so they offer higher prices to customers with higher referral prices. However if agents' estimates of referral price were equally accurate across the customer population, the proportion of sales would be constant and independent of price. The trend shown in the empirical willingness to pay curve suggests some form of bias in agents' estimates. More specifically, a lower acceptance rate for low margins implies that agents tend to overestimate the referral price of low-paying customers, so that

these customers are more likely to reject their quotes. Whereas the higher acceptance rate at higher margins implies that agents are underestimating the referral price of high-paying customers.

The exploratory data analysis in the previous section provided strong evidence for the joint dependence of price and willingness to pay on a number of covariates, including the volume of the order and the customer's purchase history. These serve to confound the relationship between price and the probability of acceptance, and must be included in the model in order to estimate the causal effect of price on willingness to pay. Estimation is complicated further by unobserved sources of dependence, in which case instrumental variables must be used to correct the resulting endogeneity bias.

## 4.2 Modelling procedure

The following section describes the steps taken to fit the joint bivariate Gaussian model from Chapter 3 to the heating oil data introduced here. The model is fitted in two stages; firstly by performing a linear regression for the seller's pricing process (the quote price model) and then a probit regression for the customer's purchasing decision (the referral price model).

### 4.2.1 Heteroscedasticity

An initial quote price model was fitted including all of the covariates in Table 4.1.1 with a significant effect (measured by BIC), and also including random effects for

`cust`, `agent`, and `day`. Diagnostics for the model revealed heteroscedasticity in the residuals. In particular, the residual variance ranges from 0.72 to 5.33 when the residuals are grouped according to depot. Figure 4.2.1 displays kernel density estimates of the residuals from the initial quote price model for each depot. The plot suggests variance is not constant across the different depots in the sample, and displays some skewness for depots with a larger spread of residuals. Levene's test also rejects the constant variance assumption (Levene, 1960). Whilst heteroscedasticity does not cause OLS estimates to be biased, it can bias standard errors and therefore affect tests of significance. This is problematic for the identification of instrumental variables, which are required to have a strong association with price.

The heteroscedasticity of residuals across depots is symptomatic of different pricing processes for depots. Given that the competitive climate and size of the local market varies considerably for each depot, it is reasonable to assume that sales agents will vary price offers according to the depot delivering the order. The difference in the distribution of residuals across depots suggests that residual variation in prices,  $\sigma_q^2$ , should not be constant over all depots. In this case, a separate quote price model and corresponding referral price model can be fitted to subsets of depots which demonstrate distinct pricing behaviour. Subsets were determined by hierarchical clustering of the residual variances for each depot, in order to identify groups of depots for which the constant variance assumption is realistic. A complete-linkage method was used, whereby the distance between two clusters is measured by the largest Euclidean distance between a member of one cluster and a member of the other. In each iteration, the pair of clusters having the shortest distance are merged (e.g. Everitt et al. (2011)).

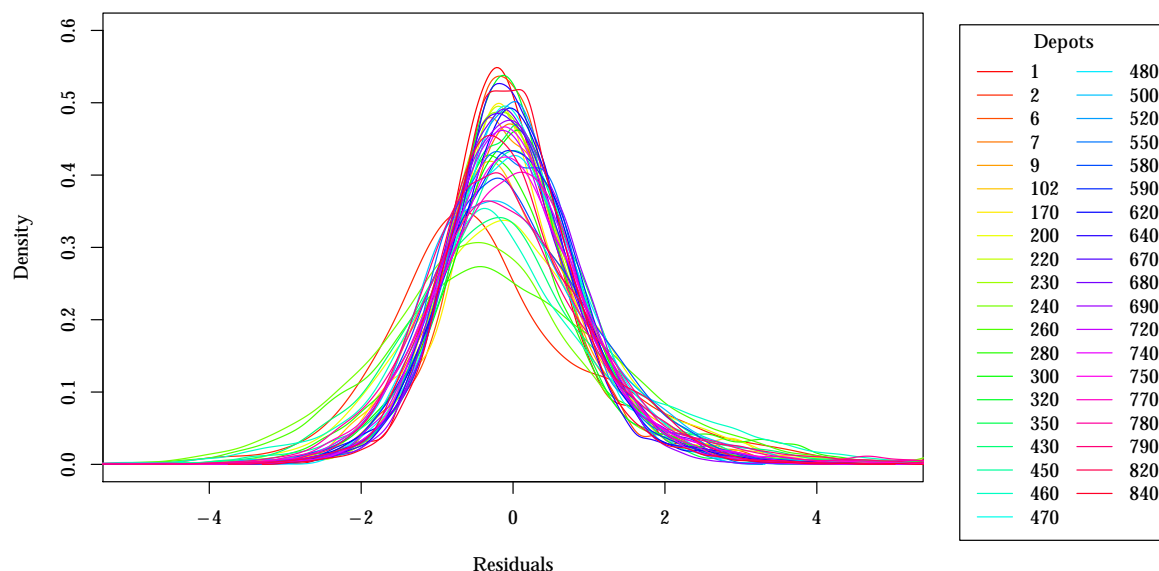


Figure 4.2.1: Kernel density plots of residuals from initial quote price model, categorised by depot.

Figure 4.2.2 shows a dendrogram of the residual variances for each depot. A suitable grouping of depots was determined using an ‘elbow plot’, in which the ratio of between group variance and total variance is plotted for different numbers of clusters. This ratio can be interpreted as the proportion of variance explained by the clusters, and as the number of clusters increases, the marginal gain - in terms of proportion of variance explained - will fall. The number of clusters is selected as the point on the elbow plot at which the marginal gain becomes negligible. On this basis, the depots were divided into three groups/subsets which are detailed in Table 4.2.1.

One potential drawback to dividing the sample and fitting a separate model to each subset is a loss of precision. The fixed effects parameters will be well-identified due to the large sample size of each subset. Random effects, however, may be affected if they consist of a large number of subjects and a small number of quotes per subject (see



Section 3.6.4). There is some overlap between the customers and agents appearing in each subset, which means that by dividing the sample into subsets, there is a loss of information in these cases. The impact on sales agent effects is minimal, since the number of observations for each agent ranges from 20 to over 6,000. In terms of customers, the vast majority (87.8%) requested quotes from the same depot and the remaining customers used two depots. Of this 12.2% of customers, 95% requested quotes from depots within the same subset. This implies only a minor loss of precision in fitting three separate models as opposed to one model for the entire dataset.

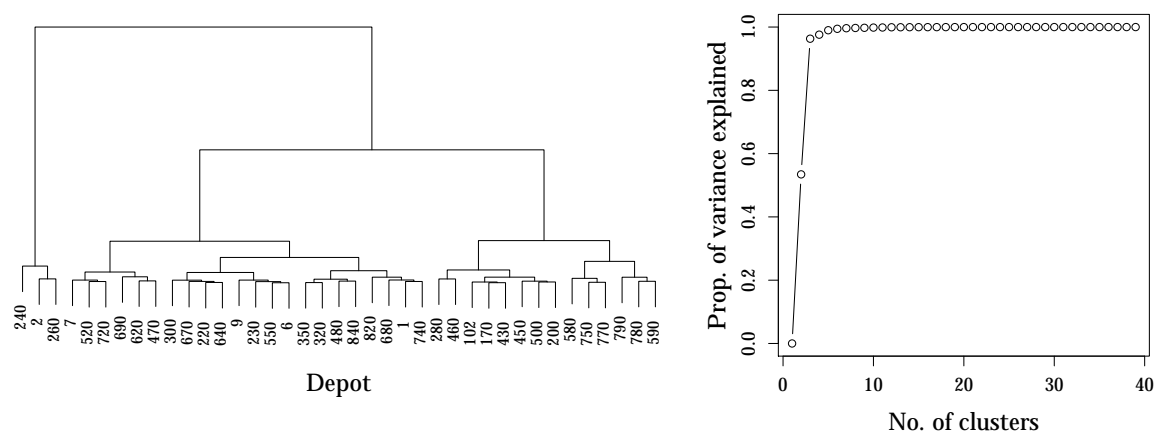


Figure 4.2.2: Hierarchical clustering of residual variances for each depot, illustrated as a dendrogram (left) and a corresponding elbow plot (right).

Table 4.2.1: Data subsets

Subset	Depots	# Depots	# Quotes	# Customers	# Agents
1	1,6,7,9,220,230,300, 320,350,470,480,520, 550,620,640,670,680, 690,720,740,820,840	22	167,833	62,871	266
2	102,170,200,280,430, 450,460,500,580,590, 750,770,780,790	14	107,748	39,568	245
3	2,240,260	3	30,483	9,834	146

### 4.2.2 Non-normality

The willingness to pay function with endogeneity correction (3.3.7) is based on an assumption of joint normality between quote price and referral price. Departures from normality will invalidate the transformation of the probit coefficients into the willingness to pay parameters (3.3.11)-(3.3.13), as well as their interpretation.

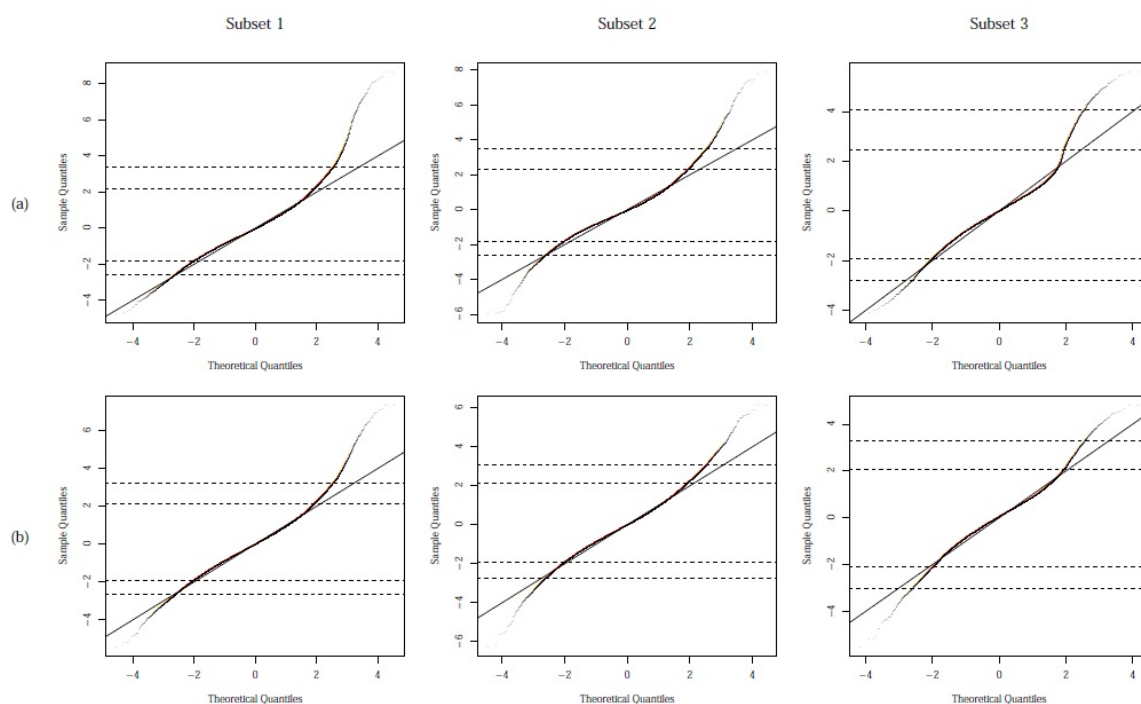


Figure 4.2.3: QQ plots of residuals from the quote price model for each subset, both (a) before and (b) after a Box-Cox transformation. The 1<sup>st</sup>, 5<sup>th</sup>, 96<sup>th</sup> and 99<sup>th</sup> sample percentiles are shown by dashed lines.

A QQ plot was used to assess the normality of the residuals,  $\hat{\epsilon}_q$ , after fitting a quote price model to each subset. The plots are shown in panel (a) of Figure 4.2.3. The residual distributions are heavily positively skewed in comparison to a normal distribution and hence a Box-Cox transformation is necessary. Maximum likelihood estimation of the Box-Cox parameter,  $\eta$ , was described in Section 3.6.1. The

profile likelihood for  $\eta$  is calculated by repeatedly fitting the model to the transformed quote prices, for different values for  $\eta$ . The estimated value of  $\eta$  for Subsets 1, 2 and 3 was 0.65, 0.4 and 0.3 respectively. QQ plots of the residuals after fitting was performed on the transformed margins, though still far from perfect, show some improvement. Each subset therefore uses a transformed variable for quote price,  $q^* = f(\text{retailPrice} - \text{spotPrice}; \hat{\eta})$  (3.6.1) and for ease of notation, we use  $q = q^*$ .

Figure 4.2.4 shows kernel density plots of the residuals, after dividing the sample into subsets and then transforming quote price before fitting. There is substantial improvement in heteroscedasticity and skewness, compared to the residual distribution for the combined model in Figure 4.2.1. The plots suggest a constant variance assumption is more realistic for the residuals after dividing into subsets and applying a Box-Cox transformation.

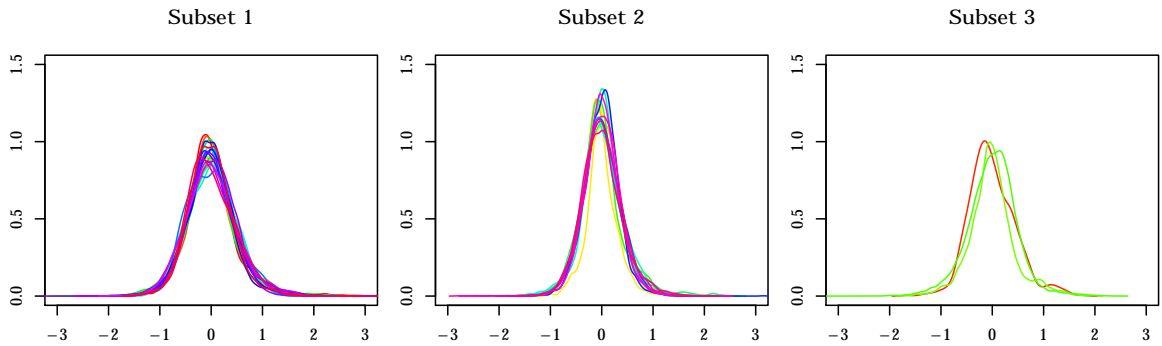


Figure 4.2.4: Kernel density plots of residuals from the final quote price model, where the sample was divided into three subsets based on depot. Residuals are grouped by depot and a colour key is provided in Figure 4.2.1.

In Section 3.6.1 it was conjectured that transformation of the price variable should have little impact on the referral price model. Since the same transformation is assumed to apply to the latent variable referral price, the goodness of fit of the model

is unchanged. A Hosmer-Lemeshow test was performed to test this notion, firstly on the predicted acceptance rates from the model with price as `margin`, and then for the transformed model with price as  $f(\text{retailPrice} - \text{spotPrice}; \hat{\eta})$  (see Section 3.6.2 for further details of the test). In both cases, the observations were divided into 100 bins and the alternative hypothesis of poor fit was rejected. Under the null hypothesis, the test statistic has a Chi-square distribution with 98 degrees of freedom. A test statistic of 91.7 and 98.2 was obtained for the untransformed and transformed models respectively, which corresponds to a p-value of 0.34 and 0.54. Figure 4.2.5 plots the observed and expected acceptance rates under each model.

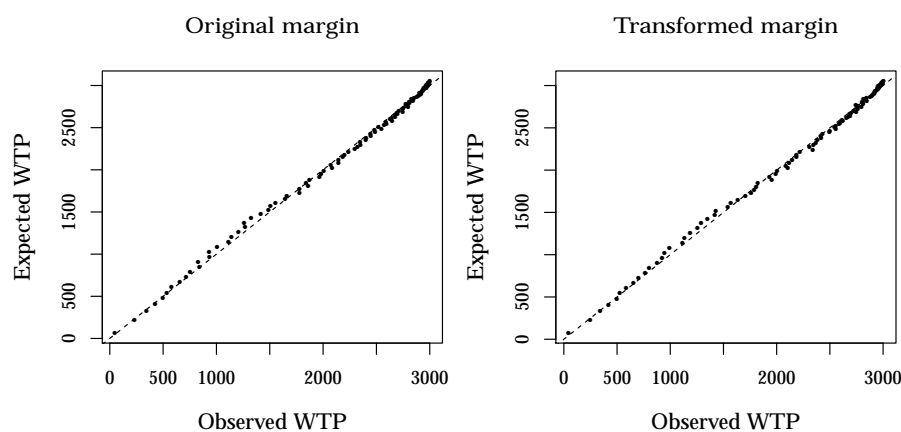


Figure 4.2.5: Observed vs. expected acceptance rates, for the WTP model with price as `margin` (left) and price as the Box-Cox transformed `margin` (right).

### 4.2.3 Model selection

In both the linear and probit regressions, a suitable model specification was determined by first selecting the fixed covariate effects. Random effects were then introduced, before re-evaluating the significance of the fixed effects. The large size of the

sample for each subset means that almost all variables are found to have a significant relationship with `margin` and `acc`, and hence automatic procedures of variable selection such as forward stepwise regression were avoided. A specific-to-general approach was employed for model selection, whereby firstly the most informative variables were identified using Bayesian Information Criterion (see Section 3.6.2). Covariates were then sequentially added and removed until a well-fitting and stable model was identified. Care was taken in noting any large changes to coefficient estimates and standard errors after the addition of new predictors, which were expected due to collinearity between temporal predictors such as `cost`, `volCount` and `temp`.

Positively skewed and large scale variables such as `dist`, `quoteInt` and `volCount`, were log-transformed. A number of different forms were considered for covariate effects, since in many cases variables were found to have a non-linear relationship with the response. Continuous explanatory variables were fitted as a piecewise-linear effect, cubic spline and also divided into bins, and then compared by BIC. The number and location of knots for the regression splines were determined by cross-validation (see Section 3.6.3 for further details). It was generally found that cubic splines provided little improvement to goodness of fit over piecewise linear effects. A price-volume effect on willingness to pay was tested by introducing an interaction term between `vol` and `margin`, and was observed to be insignificant. The number of previous responses to include in the purchase history variable, `quoteHist`, was also determined by model selection and four previous responses were found to be the most predictive.

Random effects were introduced to capture any correlation between the responses, which remained after accounting for the fixed effects. A customer, agent, day and

postcode random effect were tested for significance. In Section 3.6.4 it was mentioned that the random effects in the quote price model and referral price model need not be independent. For example, there may be a positive correlation between the agent effects for price and willingness to pay, so that agents offering higher-than-average prices are also agents with higher sales rates. An assumption of independence was validated by including the random effect in both models, and then evaluating the correlation between predictions of the effect for each model. Figure 4.2.6 displays BLUPs for the customer, agent and day random effects from each model, and shows no noticeable relationship between the effects.

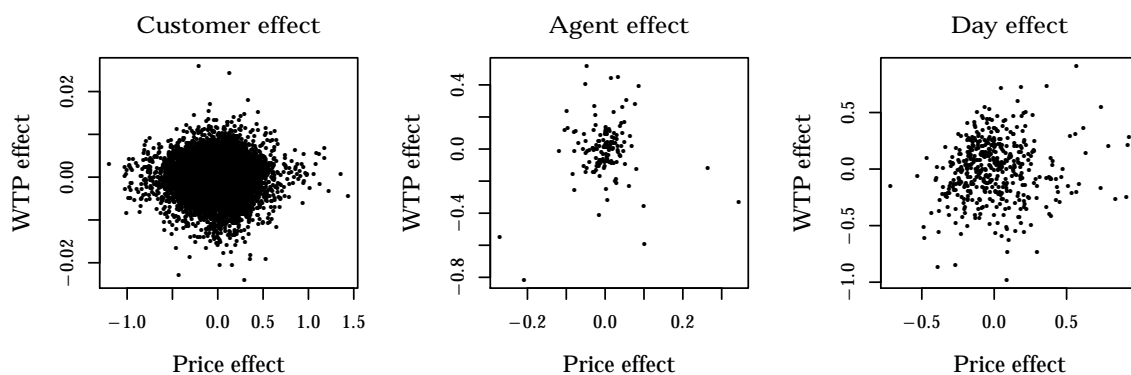


Figure 4.2.6: BLUPs of random effects in the quote price model versus effects in the referral price model.

The improvement in BIC from introducing random effects to the referral price model was not significant, and hence random effects were not included in the probit regression. This indicates that customers' responses are not correlated, for example across agents or postcodes, after accounting for fixed effects. A customer, agent and day random effect were found to be significant in the quote price model and Section 4.4.2 discusses the results.

### 4.3 Results

MLEs of the mean parameters for quote price and referral price,  $\hat{\beta}_q$  and  $\hat{\beta}_r$ , are shown in Figures 4.3.1 and 4.3.2 respectively. These include 95% confidence intervals using estimates of the standard errors.

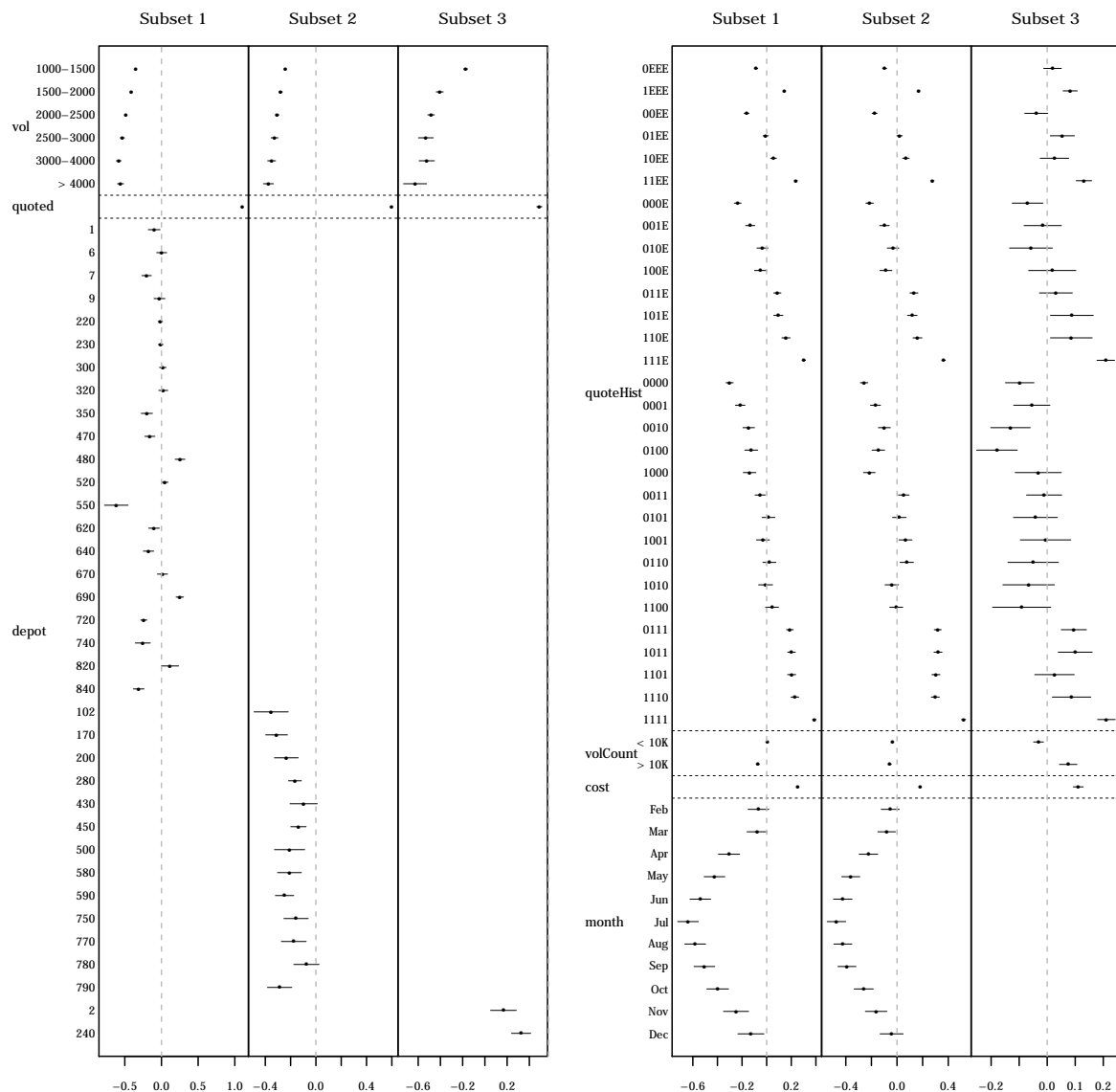


Figure 4.3.1: MLEs and 95% confidence intervals for the quote price parameters,  $\beta_q$ . The intercepts for Subsets 1-3 are 2.87, 2.72 and 2.20 respectively.

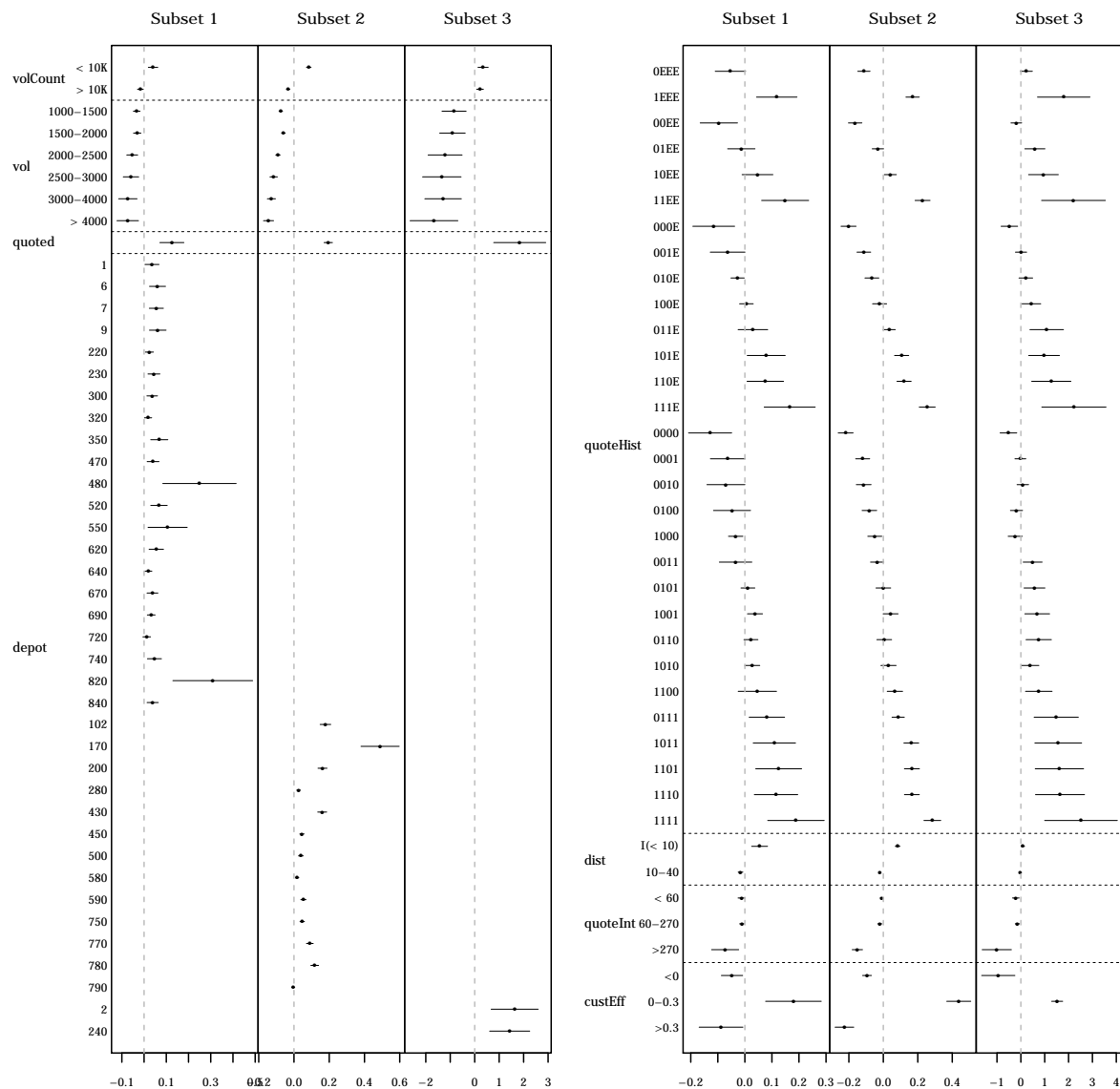


Figure 4.3.2: MLEs and 95% confidence intervals for the referral price parameters,  $\beta_r$ . The intercepts for Subsets 1-3 are -2.99, -3.17 and 0.31 respectively.

Standard errors for the referral price parameters were corrected using the bootstrap procedure outlined in Section 3.4.1. The quote price data was resampled 10,000 times and in each iteration the quote price model was refitted and the residuals were calculated; the probit regression was then performed with the new quote price residuals. The variance in the coefficients over the bootstrap sample was added to the



variance of the original probit coefficients to give the corrected standard errors. Full details of the model coefficients are provided in Tables A.0.1 and A.0.2 of Appendix A.

## 4.4 Discussion

### 4.4.1 Covariate effects

Table 4.4.1: Model contributions ( $\Delta$ BIC) of each covariate.

Covariate	Quote price model			Referral price model		
	Subset 1	Subset 2	Subset 3	Subset 1	Subset 2	Subset 3
vol	10563	3713	1293	1170	955	1133
quoteHist	6521	8004	532	24789	11453	5298
quoted	5125	1621	471	302	245	156
volCount	130	300	25	763	1752	912
depot	297	10	17	2071	1703	1187
cost	4775	1844	130	-	-	-
month	181	131	-	-	-	-
dist	-	-	-	763	395	27
quoteInt	-	-	-	425	409	239
custEff	-	-	-	686	980	496

Table 4.4.1 presents the model contributions of each covariate. Model contribution is measured in terms of change (or increase) in BIC when the covariate is excluded from the model, relative to the final model containing all of the listed covariates. The volume requested for each order and the customer's purchase history were by far the most important predictors of price. Since price here is the margin per litre, a gross volume effect was not required. Bins were created for `vol` and the price discount increases monotonically across these bins. A relatively broader range of

volume discounts is observed for Subset 3 and the volume-related parameters for referral price follow a similar pattern.

Quote price coefficients for purchase history, `quoteHist`, indicate that existing customers are offered higher prices than new customers. Furthermore, if two customers have previously requested the same number of quotes, the customer with fewer acceptances will be offered lower prices. The timing of previous acceptances appears to have a strong effect on referral price compared to quote price. For example, in the 100E and 001E categories, both customers have three previous quotes, of which one was accepted; however in the first category customers accepted their most recent quote, whilst in the second category customers accepted their oldest quote. The difference between the coefficients for these categories is relatively larger than in the referral price model compared to the quote price model. The contribution of `quoteHist` to the referral price model is substantially higher than the other covariates, which suggests that the customer's previous purchases are a strong predictor of their decision to purchase.

The variable `quoted` indicates whether the sales agent offered a price which deviated from the price recommended by the CRM system. We would expect lower variability in prices from the CRM system, however kernel density plots showed no noticeable difference in variability between price offers based on the CRM system and price offers chosen by sales agents. This may be due to differences in the variability of these quotes being overshadowed by day-to-day variation in prices. The quote price and referral price coefficients for `quoted` were both positive, which indicates that agents who offered prices based on the CRM system generally offered lower prices and

these quotes were more likely to be accepted.

Seasonality in demand is partly captured by `volCount`, which measures the volume of fuel sold at the depot on the day of the quote. The variable is logged and enters the model as a piecewise-linear effect, with a knot at 10,000 litres. The coefficient estimates suggest that during periods of high demand, where `volCount`  $> 10,000$ , prices are generally lower and willingness to pay is higher. Subset 3 is an exception to this and prices increase with sales, due to `month` not featuring in the quote price model. Since the referral price model excludes other seasonal predictors such as `month` and `cost`, `volCount` is an important predictor of willingness to pay. Likewise, a factor for depots has a substantially larger contribution to the referral price model compared to the quote price model.

Distance between the customer and depot was found to be a significant predictor of willingness to pay. An offset was added for local customers (those living less than 10km from the depot), and distances between 10 and 40km were logged and incorporated into the referral price model. A negative slope was estimated, meaning that customers living further away are less likely to accept a quote. The number of days since the customer's previous quote, `quoteInt`, was also logged and included in the referral price model as a piecewise linear effect. Knots were estimated at 60 and 270 days, or approximately 2 and 9 months. The associated parameter estimates suggest that customers with a more recent previous quote, `quoteInt`  $\leq 60$ , and customers with an older previous quote, `quoteInt`  $> 270$ , are less likely to accept a quote than those in the intermediate range. The latter result is expected as customers in this category are likely be purchasing fuel elsewhere.

As a robustness check, the models were fitted with a logit-link function and the estimated coefficients were compared to those obtained from the probit regressions. The logit coefficients were normalised by  $\pi/\sqrt{3}$  to account for the relatively higher variance of the logistic distribution (Amemiya, 1981). There were negligible differences between the coefficients, which suggests that the parameter estimates are robust to the choice of link function.

#### 4.4.2 Random effects

The customer, agent and day random effects for quote price are denoted by  $\mathbf{b}_c$ ,  $\mathbf{b}_a$  and  $\mathbf{b}_d$  respectively, with associated variances of  $\sigma_c^2$ ,  $\sigma_a^2$  and  $\sigma_d^2$ . Table 4.4.2 presents some figures related to different sources of variation in the quote price. Fixed effects account for approximately half of the variation in `margin`, as indicated by the  $R^2$  value. After the introduction of random effects,  $R^2$  increases to similar values reported in other customised pricing models (Elmaghraby et al., 2012; Phillips, 2010). Variance of the customer random effect is relatively large compared to the agent and day random effects. This suggests that differences in the prices offered to customers accounts for a large proportion of the variation in price.

Table 4.4.2: Decomposing quote price variability.

Subset	$R^2$		$\hat{\sigma}_q^2$	$\hat{\sigma}_c^2$	$\hat{\sigma}_a^2$	$\hat{\sigma}_d^2$
	Fixed effects	Mixed effects				
1	0.546	0.780	0.539	0.365	0.185	0.195
2	0.458	0.786	0.438	0.362	0.243	0.163
3	0.391	0.639	0.526	0.344	0.115	0.243

The interpretation of customer and agent random effects in the quote price model

is straightforward; the former represents systematic deviations from the average margin for each customer, whilst the latter captures any differences between the margins offered by particular agents. A random effect for the day of the quote allows the margins offered on any given day to deviate from the expected value by a consistent amount. In particular, a day effect captures day-to-day changes in the selling environment, such as limited delivery capacity, which are not captured in the data but nonetheless affect price offers. Predictions of the day effect were tested for serial correlation using the Durbin-Watson test, and were not found to be autocorrelated (Durbin and Watson, 1950). This suggests that the random effect arises out of market conditions specific to each day, rather than persistent, seasonal effects.

Figure 4.4.1 shows BLUPs of the agent price effect against the average customer effect, where the average was taken over customers requesting a quote from each agent. To check for independence between the random effects, the Pearson correlation coefficient was computed in pairwise comparisons of the BLUPs for each random effect. In all three cases the coefficient was less than 0.03, and so the random effects can be assumed to be independent of each other. In Figure 4.4.2, predictions of the customer random effect for price have been averaged by postcode. There is no obvious spatial pattern in the customer effects, which is unsurprising given that a postcode random effect was not found to be significant. However, it can be seen that customers with a high price effect tend to be isolated from other customers. This may be due to higher delivery costs for these customers.

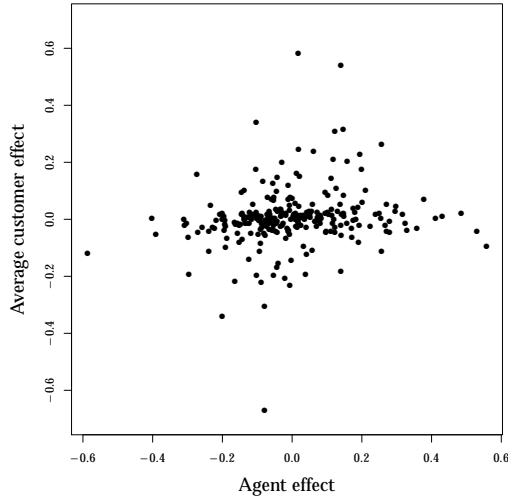


Figure 4.4.1: Average customer price effect against agent price effect.

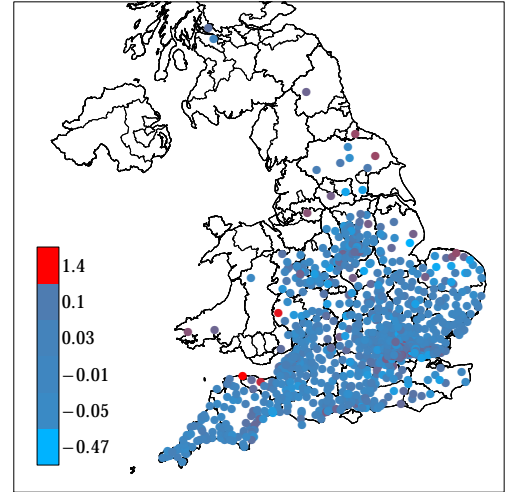


Figure 4.4.2: Average customer price effect for each postcode.

A plot of the estimated customer price random effect,  $\hat{\mathbf{b}}_c$ , against willingness to pay in Figure 4.4.3 shows a strong association, which is a manifestation of endogeneity bias. Endogeneity of price is caused by the agent's ability to identify differences in customers' willingness to pay and adjust prices accordingly. The customer price effect captures systematic differences in the prices offered to particular customers, and can therefore be interpreted as the agent's crude estimate of willingness to pay. Put another way, if endogeneity is represented by omitted variable bias, the customer price effect acts as a proxy for the omitted variable causing dependence between quote and referral price. The customer price effect, `custEff`, is therefore included in the referral price model to dampen the effects of endogeneity bias.

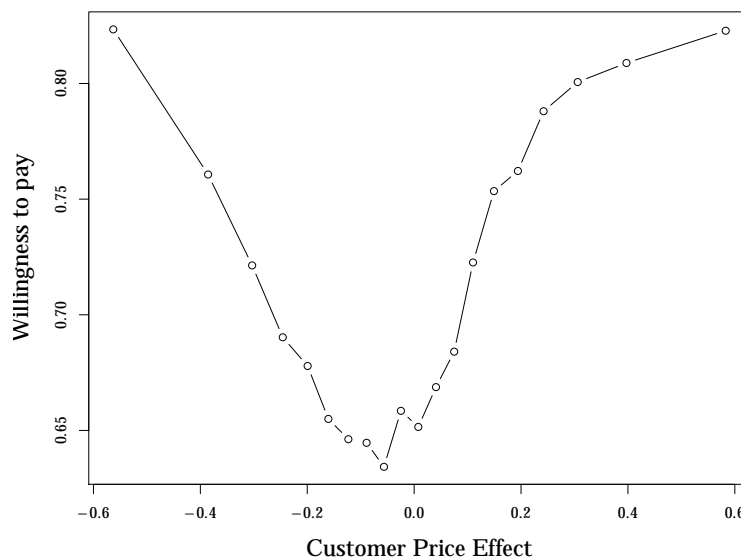


Figure 4.4.3: Relationship between the customer price effect, `custEff`, and `acc`. Observations have been binned according to `custEff`.

The probability of sale is lower for customer effects near zero, and higher for customers quoted above average ( $b_c \gg 0$ ) and below average ( $b_c \ll 0$ ) prices. These are the prices which deviate most from the expected values predicted by the quote price model. This suggests that when agents increase or decrease the price they offer for a particular quote, they are more likely to make the sale. To represent this relationship, `custEff` was included as piecewise linear effect in the referral price model with knots at 0 and 0.3.

### 4.4.3 Instrumental variables

Two instrumental variables were identified in the data: `cost` and `month`. Cost here is the difference between the wholesale price and the spot price of kerosene, and therefore represents the cost of each transaction to the seller. Customers are generally unaware

of wholesale prices and so it can be assumed that cost is independent of willingness to pay, whilst also having a strong correlation with price (see Figure 4.4.4). Month is a less obvious choice of instrumental variable and is largely effective because prices are more seasonal than willingness to pay. In other words, the month coefficients capture seasonal differences in price which remain after accounting for seasonal covariates and introducing a day random effect. Seasonal effects in the referral price model are captured by the covariate `volCount` (see Table 4.4.1), and hence `month` is not a significant predictor of willingness to pay.

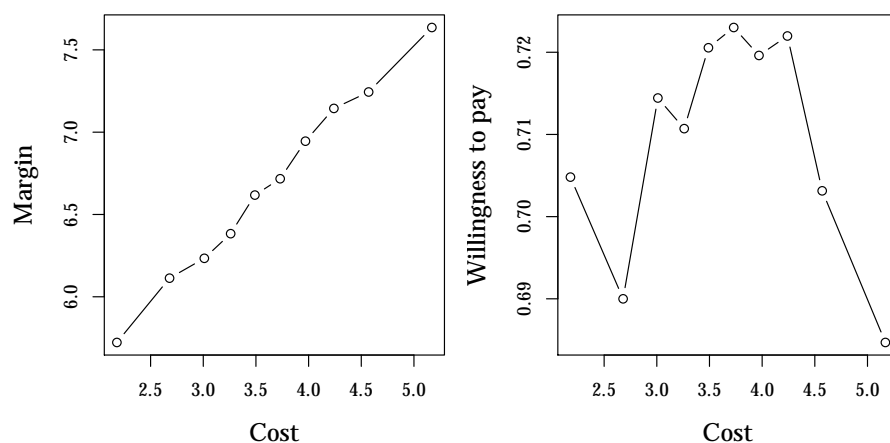


Figure 4.4.4: Relationship between cost and margin (left) and cost and observed acceptance rates (right)

In order to produce consistent estimates, instrumental variables must be independent of willingness to pay (2.2.4) and strongly correlated with price (2.2.5). These conditions are signalled by the significance of `cost` and `month` in the quote price model and their absence from the referral price model (see Table 4.4.2). Cost in particular is a strong determinant of price, as indicated by the associated  $\Delta\text{BIC}$  value in Table 4.4.1. To further demonstrate the association between the instrumental variables and



price, the quote price model was re-estimated after excluding each of the instruments in turn. Figure 4.4.5 shows the relationship between the residuals of each regression and the excluded instrument, and represents any association between margin and the instrumental variable which remains after accounting for other covariates in the model. There is clear evidence of a strong positive correlation between cost and margin. The average residual for each month is significantly different from zero, which implies that `month` is a significant predictor of margin.

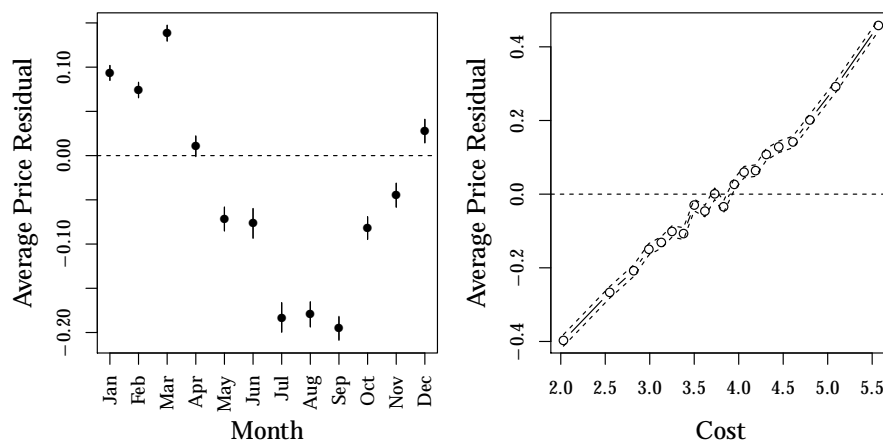


Figure 4.4.5: The instrumental variables, `month` and `cost`, were each excluded from the quote price model for Subset 1 and are plotted here against the residuals. Left: average residual for each month, with 95% confidence intervals (—). Right: average residual for different intervals of cost, with 95% confidence intervals (.....).

The independence of instrumental variables from willingness to pay is difficult to assess, since the referral price is a latent variable and the residual terms are not observable. Figure 4.4.6 shows the relationship between the instrumental variables and the difference between the observed and predicted acceptance rates from the estimated referral price model. The standard errors were calculated according to the method outlined in Section 3.4.2. Since these predictions are on the probit scale, each

prediction is normally distributed and the variance of the average prediction for a group of quotes is found by summing the variance of individual predictions. Although acceptance rates have been underestimated over June and July, overall there does not appear to be an association between the instruments and the difference between observed and predicted willingness to pay. We can therefore conclude that `month` and `cost` have little to no association with willingness to pay. Other instruments were tested, such as a lag of the spot price and day to day movements in the spot price, but were not found to be significant in the quote price model and therefore failed the relevance assumption (2.2.4).

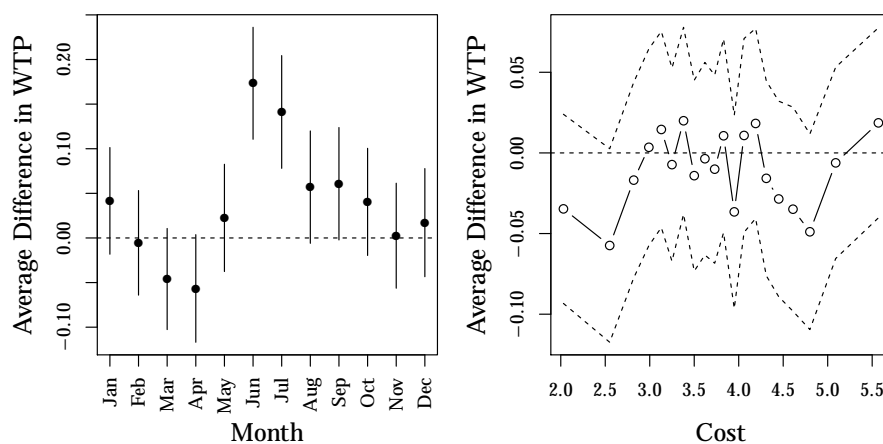


Figure 4.4.6: The difference between average and predicted acceptance rates, based on the referral price model, alongside 95% confidence intervals. Observations are grouped by month (left) and for different intervals of cost (right).

#### 4.4.4 Estimates of price sensitivity

Estimates and 95% confidence intervals for the price sensitivity parameters in the willingness to pay function are given in Table 4.4.3. The coefficients for quote price and the quote price residual,  $\lambda$  and  $\psi$ , were transformed to give the referral price

parameters,  $\sigma_r$  and  $\rho$  using (3.3.11)-(3.3.13). For all three subsets the residual correlation is significantly different from zero, which confirms the presence of endogeneity or residual correlation between price and willingness to pay. The residual correlation is highest in Subset 3, for which the quote price model also had the lowest  $R^2$  value. Since a higher proportion of unexplained variation in price equates to fewer observed sources of dependence between price and willingness to pay, a low  $R^2$  value could point to omitted variable bias.

Table 4.4.3: Williness to pay parameters

		Subset 1	Subset 2	Subset 3
$q$ coefficient	$\hat{\lambda}$	-0.128 (-0.150, -0.106)	-0.228 (-0.259, -0.197)	-1.315 (-1.377, -1.252)
$\hat{\varepsilon}_q$ coefficient	$\hat{\psi}$	0.118 (0.091, 0.145)	0.204 (0.163, 0.245)	0.537 (0.466, 0.608)
	$\hat{\sigma}_r$	7.808 (5.934, 9.682)	4.412 (3.528, 5.296)	0.790 (0.741, 0.829)
	$\hat{\rho}$	0.063 (0.041, 0.085)	0.089 (0.062, 0.116)	0.272 (0.221, 0.323)
	$\hat{\sigma}_q$	0.539 (0.535, 0.543)	0.438 (0.435, 0.442)	0.526 (0.518, 0.535)

Estimates of the referral price variance are relatively larger than associated estimates of the quote price variance in Table 4.4.2. Since higher variability in referral price corresponds to lower price-sensitivity, this indicates a high price elasticity for heating oil. In Section 3.2 it was stated that positive price elasticity is observed iff  $\rho > \sigma_q/\sigma_r$  - residual correlation is higher than the ratio of quote and referral price standard deviation. Parameter estimates show that this condition is not met by all three sets of depots, and hence price elasticity is found to be negative. Willingness to pay is estimated to decrease as prices rise, and the positive slope in the empirical

willingness to pay curve (Figure 4.1.15) is captured by the covariates and residual correlation.

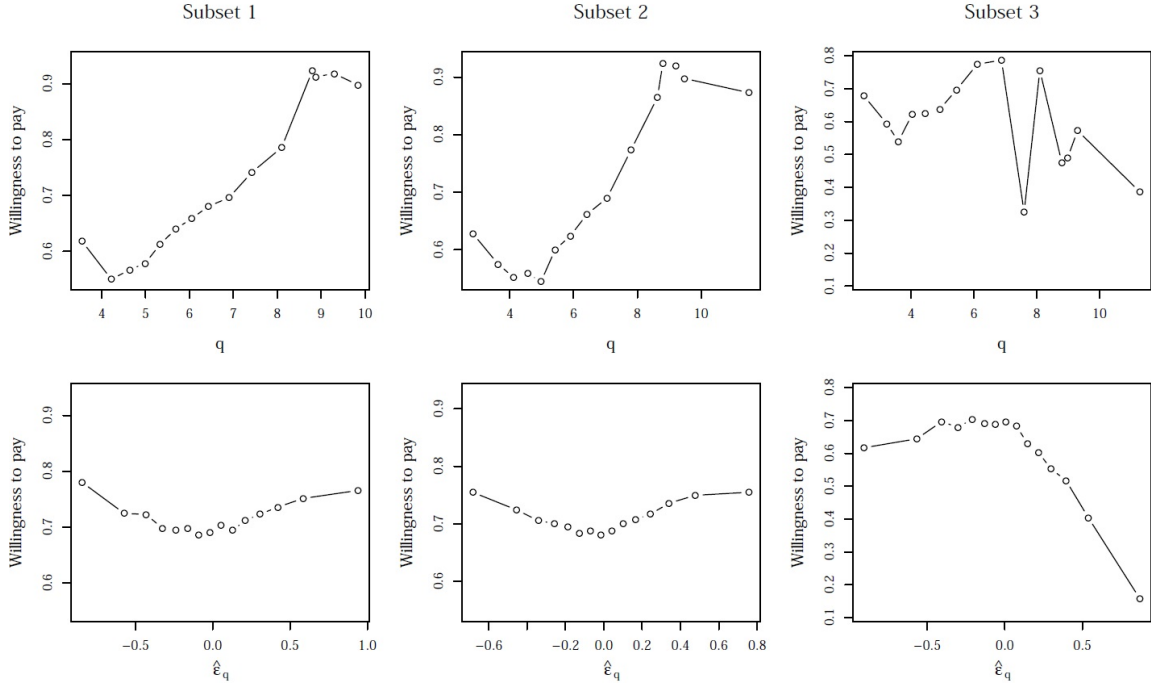


Figure 4.4.7: Empirical willingness to pay curves for each subset (top) and the relationship between quote price residuals and willingness to pay.

Recall the simulated example in Section 3.5, whereby the relationship between the quote price residual and willingness to pay was either increasing or decreasing, depending on the extent of unobserved dependence between  $q$  and  $r$ . Figure 4.4.7 shows the relationship between margin and willingness to pay, and the quote price residual and willingness to pay for each subset. Like the empirical willingness to pay curve in Section 4.1.4, Subsets 1 and 2 display an upward sloping curve and positive price elasticity. However the shape of the curve changes after accounting for joint covariate affects between the quote price and referral price,  $\mathbf{x}_c$ . Willingness to pay does not monotonically decrease with  $\varepsilon_q$ , since there is residual correlation between

the two.

In order to evaluate the degree of endogeneity in the data the standard willingness to pay function, which assumes independence between price and willingness to pay, was fitted (3.3.14). In this scenario, a quote price model is unnecessary because noise in the quote price does not affect the probability of sale. Covariate effects for the standard model were selected using BIC and in addition to the covariates previously included in the corrected referral price model, **depotCount** was found to be significant. This is consistent with the findings of Petrin and Train (2010), who note that predictors which are insignificant in the corrected model can be significant in the uncorrected model, due to a false sense of precision in assuming price is independent of unobserved factors.

Table 4.4.4: Willingness to pay parameters under independence

	Subset 1	Subset 2	Subset 3
$q$ coefficient	0.014 (0.004, 0.024)	0.040 (0.026, 0.054)	-0.613 (-0.642, -0.584)
$\hat{\sigma}_r$	-73.9	-25.0	1.63

The estimated coefficient for price for each subset is shown in Table 4.4.4, which corresponds to  $\lambda$  in the corrected referral price model. The price coefficients for Subsets 1 and 2 are positive, indicating a positive relationship between price and willingness to pay. In terms of referral price parameters, this corresponds to a negative and invalid estimate for referral price standard deviation. These results demonstrate a severe endogeneity bias in willingness to pay estimates, which is corrected by the model developed here.

## 4.5 Summary

In this chapter, the model presented in Chapter 3 was developed to represent the pricing and purchasing processes involved in the sale of heating oil. Prices are highly variable within this market due to customer heterogeneity, volatile costs and seasonality of demand. Positive price elasticity is observed in the sales data, whereby the proportion of acceptances increases with price. This occurs because sales agents are able to vary prices on a transaction-to-transaction basis, according to their estimate of the customer's willingness to pay.

An appropriate model was first developed for the agents' pricing process, using the available covariates. Initial results suggested that pricing behaviour is not homogenous across depots and an assumption of normality for residual variation in price is not realistic. The depots were therefore divided into three subsets and a separate model was fitted to each subset, after applying a Box-Cox transformation to price. Customer, agent and day random effects for price were included in the quote price model to account for residual correlation between price offers. Two instrumental variables were identified, month and cost, which were found to be significant predictors of price and approximately independent of willingness to pay.

The residuals from the pricing model were then entered into a probit regression of the customer responses, alongside predictions of the customer price effect. This had the effect of isolating variation in willingness to pay caused directly by price, and hence provided estimates of true price sensitivity. For all three subsets, parameter estimates returned negative price elasticity and since  $\hat{\rho} \neq 0$ , pointed to an unobserved source

of dependence between quote price and referral price. This was confirmed by naive estimation of willingness to pay, in which prices were assumed to be independent, and invalid estimates of the referral price parameters were obtained.

# Chapter 5

## Profit simulation

In this chapter, the impact of price changes on profit are evaluated by Bayesian estimation of willingness to pay. The analysis is carried out for Subset 3, as it has the fewest quotes and therefore the lowest computational cost. However in the absence of time constraints, the same process could be performed on Subsets 1 and 2. The model introduced in Chapter 3.1 is fitted by an MCMC scheme and the resulting sample of parameters is used to predict expected profit under a number of price change scenarios.

### 5.1 Introduction

In Chapter 2 the unconstrained price optimisation problem was introduced:

$$\dot{q} = \operatorname{argmax}_q h(q, c, V) \times \pi(q; \mathbf{x}). \quad (5.1.1)$$

Given available information about a customer and their order,  $\mathbf{x}$ , the price,  $q$ , offered to the customer can be selected to maximise the expected profit from the



sale. This is estimated by the product of the profit earned by the seller if the quote is accepted,  $h(q, c, V)$ , and the customer's probability of accepting the quote,  $\pi(q; \mathbf{x})$ . The latter is referred to as willingness to pay. An automated price recommendation system could perform this operation for millions of transactions in real-time, enabling the seller to optimise prices for each individual transaction. In order to maintain accuracy, predictions of willingness to pay would be monitored and kept in line with observed acceptance rates by refitting the model.

The subject of this chapter is a counterfactual analysis, in which the prices of historical quotes are modified and their impact on expected profit is examined to provide insights into customers' purchasing behaviour under different price scenarios. In this way, we can evaluate different features of the model by considering which covariates are most influential to willingness to pay, and thereby identify opportunities to improve profit. The profit earned from existing prices is compared to the expected profit that would have been earned from alternative prices. We investigate small price changes rather than optimal prices for two reasons: firstly, it is crucial that any predictions are made within the range of validity of the model and secondly, a seller would be more willing to implement changes with a relatively low impact on profit. The latter is especially true when, as in this case, pricing authority is granted to sales agents rather than prices being set centrally.

A Bayesian approach to the analysis is preferred in order to quantify our uncertainty in predictions of profit. In the previous chapter, confidence intervals for parameters were based on standard error estimates and the asymptotic normality of maximum likelihood estimators. Combining the standard errors of each MLE into

an overall measure of uncertainty for profit is not straightforward, and simply taking point estimates to predict profit does not properly account for uncertainty in the prediction.

Bayesian estimation was introduced in Section 2.4, and is based on the posterior distribution. For intractable problems such as the one examined here, MCMC simulation is performed to obtain a sample of parameters from the posterior. This posterior sample is then used to approximate the posterior predictive distribution for a customer's response to a new price quote, which in turn provides a posterior sample for profit. From this posterior sample, we can predict the profit earned by the seller under a given set of prices. Ideally, predictions would account for the uncertainty of all parameters by sampling from the full posterior of the model. However, some parameters are known with high precision from the likelihood-based fit of the quote price model and so, to reduce the cost of the analysis, Bayesian estimation is only performed on the remaining parameters (see Section 5.2.1 for further details).

It is important to account for parameter uncertainty when evaluating the quality of different pricing policies, since the seller offers prices to thousands of customers over the course of a day and even minor price changes can have far-reaching effects. We wish to avoid scenarios in which we have low confidence in the predicted impact of a price change, and in particular, scenarios where the price change has a significant probability of negatively affecting profit.

The sections below define the probability of sale for an existing quote under a new price, and outline how profit is calculated and compared. This is followed by a recap of the willingness to pay model, a description of the Bayesian estimation procedure,

and an evaluation of the parameter estimates. The proposed price changes are then introduced and the estimated profit improvement under each policy is presented and discussed.

### 5.1.1 Predicting willingness to pay

For each quote in the sample we have observed the price quoted to the customer,  $q$  and their purchase decision  $y$ , as well as a number of related covariates,  $\mathbf{x}$ . We now consider an alternative price,  $\tilde{q}$ , and associated purchase decision,  $\tilde{Y}$  for each quote. Predictions of  $\tilde{Y}$  are made from the posterior predictive distribution:

$$p(\tilde{y}|\tilde{q}, y, q) = \int p(\tilde{y}|\tilde{q}, y, q, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{q}) \, \mathbf{d}\boldsymbol{\theta}. \quad (5.1.2)$$

This is the expectation of the likelihood,  $p(\tilde{y}|\tilde{q}, y, q, \boldsymbol{\theta})$ , of  $\tilde{y}$  over the posterior distribution,  $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{q})$ . In other words, this is the probability distribution of the customer's response to  $\tilde{q}$ , given the data already observed, after averaging over the willingness to pay parameters. Then given a posterior sample of the parameters, we can obtain a Monte Carlo estimate of (5.1.2), as described in Section 2.4. For this we require an expression for the probability of sale at an alternative quote price of  $\tilde{q}$ , conditional on the customer's response to price  $q$ .

We have that  $\tilde{Y}|\tilde{q}, y, q \sim \text{Bernoulli}(\tilde{\pi}(\tilde{q}; y, q, \mathbf{x}, \boldsymbol{\theta}))$ , where:

$$\tilde{Y} = \begin{cases} 1 & \text{if } R > \tilde{q} \text{ i.e. } \tilde{q} \text{ is accepted} \\ 0 & \text{otherwise.} \end{cases} \quad (5.1.3)$$

The customer's referral price distribution is conditional on the price offered by the agent,  $q$  and the covariates,  $\mathbf{x}$  (as shown in Section 3.1). Before learning the customer's response, the probability that the customer's referral price exceeds the alternative price  $\tilde{q}$  is given by:

$$P(R > \tilde{q} | q, \mathbf{x}) = \Phi \left( \frac{1}{\sqrt{1 - \rho^2}} \left[ \frac{1}{\sigma_r} \mathbf{x}'_r \boldsymbol{\beta}_r - \frac{1}{\sigma_r} \tilde{q} + \frac{\rho}{\sigma_q} q - \frac{\rho}{\sigma_q} \mathbf{x}'_q \boldsymbol{\beta}_q \right] \right). \quad (5.1.4)$$

And correspondingly,

$$\begin{aligned} P(R \leq \tilde{q} | q, \mathbf{x}) &= 1 - P(R > \tilde{q} | q, \mathbf{x}) \\ &= \Phi \left( -\frac{1}{\sqrt{1 - \rho^2}} \left[ \frac{1}{\sigma_r} \mathbf{x}'_r \boldsymbol{\beta}_r - \frac{1}{\sigma_r} \tilde{q} + \frac{\rho}{\sigma_q} q - \frac{\rho}{\sigma_q} \mathbf{x}'_q \boldsymbol{\beta}_q \right] \right). \end{aligned} \quad (5.1.5)$$

After observing  $y$ , the referral price distribution becomes a truncated Normal and the probability of sale at an alternative price is found from Bayes' theorem. The equations in the remainder of this section are implicitly conditional on the specific values of  $q$  and  $\tilde{q}$ . Suppose that the customer accepted a quote of  $q$ . Then the probability of the customer accepting an increased price of  $\tilde{q}$  is given by:

$$\begin{aligned} P(\tilde{Y} = 1 | \tilde{q} > q, y = 1, \mathbf{x}) &= P(R > \tilde{q} | \tilde{q} > q, y = 1, \mathbf{x}) \\ &= \frac{P(R > \tilde{q}, R > q | \tilde{q} > q, \mathbf{x})}{P(R > q | \mathbf{x})} \\ &= \frac{P(R > \tilde{q} | \mathbf{x})}{P(R > q | \mathbf{x})}. \end{aligned} \quad (5.1.6)$$

Under a price increase, quotes which were rejected at  $q$  will also be rejected at  $\tilde{q}$ ,

since:

$$\begin{aligned} \mathrm{P}\left(\tilde{Y} = 1 \mid \tilde{q} > q, y = 0, \mathbf{x}\right) &= \frac{\mathrm{P}(R > \tilde{q}, R \leq q \mid \tilde{q} > q, \mathbf{x})}{\mathrm{P}(R \leq q \mid \mathbf{x})} \\ &= 0. \end{aligned} \tag{5.1.7}$$

Similarly, under a price decrease, quotes which were accepted at the original price will also be accepted at the new price:

$$\begin{aligned} \mathrm{P}\left(\tilde{Y} = 1 \mid \tilde{q} < q, y = 1, \mathbf{x}\right) &= \frac{\mathrm{P}(R > \tilde{q}, R > q \mid \tilde{q} < q, \mathbf{x})}{\mathrm{P}(R > q \mid \mathbf{x})} \\ &= \frac{\mathrm{P}(R > q \mid \mathbf{x})}{\mathrm{P}(R > q \mid \mathbf{x})} \\ &= 1. \end{aligned} \tag{5.1.8}$$

The probability that  $\tilde{q}$  will be accepted by the customer given that the higher price of  $q$  was rejected is given by:

$$\begin{aligned} \mathrm{P}\left(\tilde{Y} = 1 \mid \tilde{q} < q, y = 0, \mathbf{x}\right) &= \frac{\mathrm{P}(R > \tilde{q}, R \leq q \mid \tilde{q} < q, \mathbf{x})}{\mathrm{P}(R \leq q \mid \mathbf{x})} \\ &= \frac{\mathrm{P}(\tilde{q} < R \leq q \mid \mathbf{x})}{\mathrm{P}(R \leq q \mid \mathbf{x})} \\ &= \frac{\mathrm{P}(R \leq q \mid \mathbf{x}) - \mathrm{P}(R \leq \tilde{q} \mid \mathbf{x})}{\mathrm{P}(R \leq q \mid \mathbf{x})}. \end{aligned} \tag{5.1.9}$$

Willingness to pay for an alternative price is defined by combining equations (5.1.6)-(5.1.9):

$$\tilde{\pi}(\tilde{q}; y, q, \mathbf{x}, \boldsymbol{\theta}) = \begin{cases} \frac{P(R > \tilde{q} | \mathbf{x})}{P(R > q | \mathbf{x})} & \text{if } \tilde{q} > q, y = 1, \\ 0 & \text{if } \tilde{q} > q, y = 0, \\ 1 & \text{if } \tilde{q} < q, y = 1, \\ \frac{P(R \leq q | \mathbf{x}) - P(R \leq \tilde{q} | \mathbf{x})}{P(R \leq q | \mathbf{x})} & \text{if } \tilde{q} < q, y = 0. \end{cases} \quad (5.1.10)$$

For each set of parameter values sampled from the posterior distribution, willingness to pay is evaluated and then in turn used to sample a value of  $\tilde{Y}$  from the posterior predictive distribution using (5.1.3).

### 5.1.2 Measuring profit improvement

In order to predict profit we require the function  $h(q, c, V)$ , which returns the profit available from a sale. The term ‘available’ is used here, since zero profit is earned if the quote is rejected. In this case, available profit is given by the per litre profit, `retailPrice - wholePrice`, multiplied by `vol`. Recall that in Section 4.2.2, for modelling purposes, the price variable was selected as the Box-Cox transformed difference between `retailPrice` and `spotPrice`:

$$q = \frac{(r - s)^{0.3} - 1}{0.3}. \quad (5.1.11)$$

Here  $r$  is the retail price offered for the quote and  $s$  is the spot price at the time of the quote. Then, if  $c$  is used to denote the wholesale price and  $V$  denotes the volume of the order, the available profit from the sale is given by:

$$h(q, c, V; s) = V \cdot [(0.3 q + 1)^{1/0.3} + s - c]. \quad (5.1.12)$$

For ease of notation, the function is simplified to  $h(q)$ . The posterior predictive distribution for the actual profit earned from quoting  $\tilde{q}$  is analogous to that of  $\tilde{Y}$ , with the distinction that profit takes values of 0 and  $h(\tilde{q})$  rather than 0 and 1. Profit from the quote is therefore predicted by multiplying the posterior sample for  $\tilde{Y}$  by  $h(\tilde{q})$ .

The total profit earned from all quotes at a new set of prices,  $\tilde{\mathbf{q}}$  is denoted by:

$$W(\tilde{\mathbf{q}}) = \sum_{i=1}^N h(\tilde{q}_i) \cdot \tilde{y}_i. \quad (5.1.13)$$

This is compared to the actual profit earned by the seller under the original prices:

$$W(\mathbf{q}) = \sum_{i=1}^N h(q_i) \cdot y_i. \quad (5.1.14)$$

The impact of the price change is measured by the difference between (5.1.13) and (5.1.14):

$$\Delta W(\tilde{\mathbf{q}}, \mathbf{q}) = \sum_{i=1}^N h(\tilde{q}_i) \cdot \tilde{y}_i - h(q_i) \cdot y_i. \quad (5.1.15)$$

Profit under a new set of prices (5.1.13) is subject to both parameter uncertainty and random variation in each customer's response to  $\tilde{q}_i$ . The remainder of this chapter outlines a method for averaging over both sources of uncertainty in our estimation of (5.1.15). Firstly by simulating parameters from the posterior distribution, and then by simulating customer responses from the posterior predictive distribution. The next

section provides an overview of the Bayesian fitting process for the model.

## 5.2 Bayesian estimation

### 5.2.1 Recap of model

The full willingness to pay model from the previous chapter can be expressed as follows:

$$\mathbf{q} = \mathbf{X}_q \boldsymbol{\beta}_q + \mathbf{Z}_c \mathbf{b}_c + \mathbf{Z}_a \mathbf{b}_a + \mathbf{Z}_d \mathbf{b}_d + \boldsymbol{\varepsilon}_q, \quad (5.2.1)$$

$$\pi(\mathbf{q}) = \Phi(\mathbf{X}_r \boldsymbol{\gamma} + \lambda \mathbf{q} + \psi \boldsymbol{\varepsilon}_q). \quad (5.2.2)$$

where  $\mathbf{b}_c \sim \text{MVN}(0, \sigma_c^2 \mathbb{I}_{n_c})$ ,  $\mathbf{b}_a \sim \text{MVN}(0, \sigma_a^2 \mathbb{I}_{n_a})$ ,  $\mathbf{b}_d \sim \text{MVN}(0, \sigma_d^2 \mathbb{I}_{n_d})$  and  $\boldsymbol{\varepsilon}_q \sim \text{MVN}(0, \sigma_q^2 \mathbb{I}_n)$ . In the quote price model (5.2.1),  $\mathbf{b}_c$ ,  $\mathbf{b}_a$  and  $\mathbf{b}_d$  are the customer, agent and day random effects respectively, and  $\boldsymbol{\beta}_q$  is a  $(p_q \times 1)$  vector of fixed effect parameters.  $\mathbf{Z}_c$  is an  $(n \times n_c)$  matrix comprised of 0's and 1's, where  $n_c$  is the number of customers in the sample and each row contains a single non-zero entry indicating which customer requested the quote. The agent matrix,  $\mathbf{Z}_a$ , and day matrix,  $\mathbf{Z}_d$  are similarly defined. The instrumental variable, **cost**, is included as column of  $\mathbf{X}_q$  and is linearly independent of  $\mathbf{X}_r$ . The final two columns of  $\mathbf{X}_r$  are a piecewise linear transformation of  $\mathbf{b}_c$ , the customer random effect.

The referral price model has  $p_r + 2$  estimable parameters,  $\boldsymbol{\phi}_r = \{\boldsymbol{\gamma}, \lambda, \psi\}$ ; the  $p_r$  covariate effects,  $\boldsymbol{\gamma}$ , and the price sensitivity parameters,  $\lambda$  and  $\psi$ . The original mean and variance parameters of the referral price distribution,  $\boldsymbol{\theta}_r = \{\boldsymbol{\beta}_r, \lambda, \psi\}$ , can be



obtained by transforming these coefficients according to (3.3.11)-(3.3.13) in Chapter 3. While the choice of parameterisation does not affect predictions of willingness to pay and profit, since the posterior predictive distribution marginalises over parameters, it can impact performance of the MCMC algorithm. MCMC is performed on the raw versions of the willingness to pay coefficients,  $\phi_r$ , rather than  $\theta_r$ , since this simplifies the specification of priors and enables block updating of the parameters.

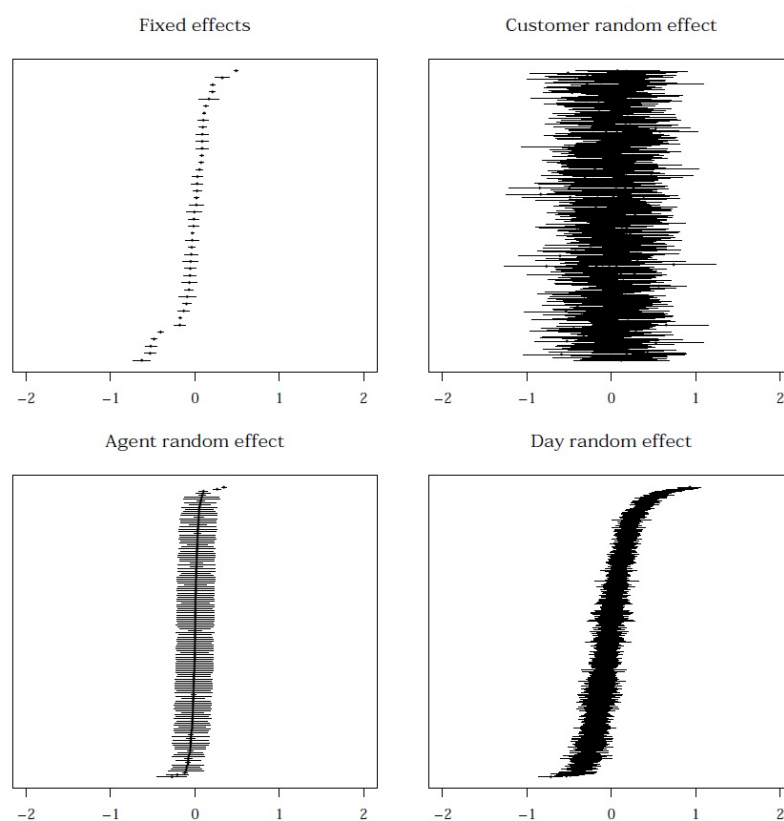


Figure 5.2.1: Point estimates and standard errors for fixed effects parameters and random effects from the quote price model for Subset 3.

Figure 5.2.1 shows point estimates and confidence intervals for the fixed effect parameters in the quote price model, and also Best Linear Unbiased Predictors (BLUPs) and prediction intervals for the random effects. Confidence intervals for the fixed ef-

fects are based on the asymptotic distribution of the MLEs, whilst prediction intervals for the random effects are based on standard errors of the BLUPs.

The fixed effects parameters are precisely estimated as they use the entire sample of 30,483 quotes. The average number of quotes each day was 67, with the least busy day having 8 quotes and hence prediction intervals of each day effect are relatively narrow. Similarly, agents make an average of 209 quotes over the study period and so the agent random effects are also well-identified. However for approximately one quarter of customers in Subset 3, only one quote is requested during the study period and the resulting estimates of customer random effects are less precise. Since the customer random effect estimates are included as a fixed effect in the willingness to pay function, this has a direct impact on the referral price parameters.

Table 5.2.1: Numbers of random effects and parameters in WTP model for Subset 3

Quote price model				Referral price model
# Fixed effects	Random effects			# Fixed effects
	# Customers	# Agents	# Days	
$p_q$	$n_c$	$n_a$	$n_d$	$p_r + 2$
43	9,834	146	455	51

Table 5.2.1 shows the number of fixed effects parameters and random effects in the full model for Subset 3. The random effects are non-nested and so the quotes associated with a particular customer may also be associated with one or two sales agents. Hence in a Gibbs sampling scheme where updates of one random effect would be conditional on the values of the other random effects, the chain would move very slowly. For these reasons, rather than the full model, Bayesian estimation is performed

on the customer random effect for price,  $\mathbf{b}_c$ , and referral price parameters,  $\phi_r$ . The estimates are conditional on the quote price parameters,  $\theta_q$  and day and agent random effects for price,  $\mathbf{b}_d$  and  $\mathbf{b}_a$ , which are fixed at their maximum likelihood estimates and BLUPs (see Sections 4.4.1 and 4.4.2 ).

Ideally, predictions of profit would take into account the uncertainty in all parameters and random effects, by sampling from the posterior distribution of the full model, but we must compromise in order to achieve this in the computational time frame available.

### 5.2.2 Prior and posterior distributions

Since MCMC estimation is performed on the raw referral price parameters,  $\phi_r$ , for the remainder of this chapter, the willingness to pay function is parameterised as follows:

$$\pi(q; \mathbf{x}_r, \phi_r) = \Phi \left( \mathbf{x}_r' \boldsymbol{\gamma} + \lambda q + \psi \hat{\varepsilon}_q \right). \quad (5.2.3)$$

The term  $\hat{\varepsilon}_q$  is used for the quote price residual rather than  $\varepsilon_q$  to indicate that the function is conditional on the quote price parameters. To make explicit the inclusion of the quote price customer effect,  $\mathbf{b}_c$ , as a piecewise linear fixed effect for referral price, we can write:

$$\pi(q; \dot{\mathbf{x}}_r, \mathbf{b}_c, \phi_r) = \Phi \left( \dot{\mathbf{x}}_r' \boldsymbol{\gamma}_1 + \left[ k \left( \mathbf{z}_c' \mathbf{b}_c \right) \right]' \boldsymbol{\gamma}_2 + \lambda q + \psi \hat{\varepsilon}_q \right). \quad (5.2.4)$$

where  $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2\}$  and  $\dot{\mathbf{x}}_r$  represents the referral price covariates not related to the customer random effect for price. In order to produce a piecewise linear effect with a

knot at zero, the function  $k(\cdot)$  transforms the customer effect into a vector containing an interaction term with a dummy variable:

$$k(x) = \begin{pmatrix} x \\ x \cdot \mathbb{I}[x > 0] \end{pmatrix}.$$

The likelihood of the referral price parameters and customer price effect, conditional on the quote price parameters and day and agent random effects for price, is given by:

$$L(\phi_r, \mathbf{b}_c | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q, \mathbf{b}_a, \mathbf{b}_d) = \sum_{i=1}^n \pi(q_i; \dot{\mathbf{x}}_{r,i}, \mathbf{b}_c, \phi_r)^{y_i} (1 - \pi(q_i; \dot{\mathbf{x}}_{r,i}, \mathbf{b}_c, \phi_r))^{(1-y_i)}.$$

A natural prior for the customer random effects in (5.2.2) is the conditional distribution of the customer effect, given maximum likelihood estimates of the quote price model. That is, the distribution of the customer random effects after having observed the prices  $\mathbf{q}$ , conditional on the other random effects and quote price parameters. Let  $C_j$  denote the set of indices corresponding to quotes from customer  $j$ , i.e.  $C_j$  denotes the set of indices of the non-zero entries in the  $j^{th}$  column of  $\mathbf{Z}_c$ . Then the prior is given by:

$$b_{c,j} | \mathbf{q}, \boldsymbol{\theta}_q \sim N \left( \frac{\sigma_c^2}{\sigma_q^2 + n_j \sigma_c^2} \sum_{i \in C_j} (q_i - \hat{\mu}_i), \frac{\sigma_c^2 \sigma_q^2}{\sigma_q^2 + n_j \sigma_c^2} \right), \quad (5.2.5)$$

where  $\hat{\mu}_{q,i} = \mathbf{x}'_{q,i} \hat{\boldsymbol{\beta}}_q + \mathbf{z}'_{a,i} \hat{\mathbf{b}}_a + \mathbf{z}'_{d,i} \hat{\mathbf{b}}_d$ , and  $\mathbf{x}_{q,i}$ ,  $\mathbf{z}_{a,i}$  and  $\mathbf{z}_{d,i}$  are the  $i^{th}$  rows of  $\mathbf{X}_q$ ,  $\mathbf{Z}_a$  and  $\mathbf{Z}_d$  respectively. A vague prior is specified for  $\phi_r$ , so that we can draw objective inferences about willingness to pay based primarily on the observed data:

$$\boldsymbol{\phi}_r \sim \text{MVN} \left( 0, \delta^2 \mathbf{I}_{p_r+2, p_r+2} \right). \quad (5.2.6)$$

$\delta$  is set as  $\sqrt{5}$ , which is approximately 8 times the size of the largest standard error estimate from maximum likelihood estimation of  $\boldsymbol{\phi}_r$ .

The joint posterior is given by the product of the priors for  $\boldsymbol{\phi}_r$  and each  $b_{c,j}$  ( $j = 1, \dots, n_c$ ), as well as the likelihood:

$$p(\boldsymbol{\phi}_r, \mathbf{b}_c | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q) \propto L(\boldsymbol{\phi}_r, \mathbf{b}_c | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q, \mathbf{b}_a, \mathbf{b}_d) p(\boldsymbol{\phi}_r) \prod_{j=1}^{n_c} p(b_{c,j} | \mathbf{q}, \boldsymbol{\theta}_q). \quad (5.2.7)$$

The conditional posteriors are then:

$$p(\boldsymbol{\phi}_r | \mathbf{b}_c, \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q) \propto L(\boldsymbol{\phi}_r, \mathbf{b}_c | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q, \mathbf{b}_a, \mathbf{b}_d) p(\boldsymbol{\phi}_r), \quad (5.2.8)$$

$$p(b_{c,j} | \boldsymbol{\phi}_r, \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q) \propto L(\boldsymbol{\phi}_r, \mathbf{b}_c | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q, \mathbf{b}_a, \mathbf{b}_d) p(b_{c,j} | \mathbf{q}, \boldsymbol{\theta}_q). \quad (5.2.9)$$

Since both of these distributions are analytically intractable, posterior samples are obtained from a random walk Metropolis within Gibbs algorithm. The method is outlined in the next section.

### 5.2.3 MCMC scheme

Gibbs sampling is performed in two stages. In the first stage, the referral price parameters are updated, conditional on the current values of the customer random effects. The customer random effects are then updated, conditional on the newly updated values of the referral price parameters. The random effect for each customer

is independent from the effects for other customers, and hence each random effect can be updated separately. A random walk Metropolis (RWM) step is used to perform updates of the customer random effects (e.g. Gilks et al. (1995)). For each iteration of the chain,  $m = 1, \dots, M$ , jumps are proposed as follows:

$$b_{c,j}^{(m)*} \sim N\left(b_{c,j}^{(m-1)}, 2.4^2 \sigma^2\right) \quad \text{for } j = 1, \dots, n_c.$$

These jumps are accepted with probability:

$$\alpha\left(b_{c,j}^{(m-1)}, b_{c,j}^{(m)*}\right) = \min\left(1, \frac{p\left(b_{c,j}^{(m)*} | \phi_r^{(m)}, \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q\right)}{p\left(b_{c,j}^{(m-1)} | \phi_r^{(m)}, \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q\right)}\right).$$

Based on some trial runs of the chain,  $\sigma^2$  was set to 0.05 in order to achieve an acceptance rate of 0.44, the optimal acceptance rate for a univariate Gaussian target distribution with a Gaussian proposal (Roberts and Rosenthal, 2001).

For high-dimensional distributions, the tuning of the variance of the proposal distribution is crucial to achieve efficient mixing. We therefore use the *adaptive* RWM algorithm described in Sherlock et al. (2010) to update the referral price parameters. Rather than having to use trial and error to tune the size of proposal jumps, adaptive MCMC learns the shape of the target distribution and scales the proposal distribution accordingly. Proposals are made from a mixture distribution:

$$\phi_r^{(m)*} \sim \begin{cases} \text{MVN}\left(\phi_r^{(m-1)}, \lambda^{(0)2} \boldsymbol{\Sigma}^{(0)}\right) & \text{w.p. } \epsilon \\ \text{MVN}\left(\phi_r^{(m-1)}, \lambda^{(m-1)2} \boldsymbol{\Sigma}^{(m-1)}\right) & \text{w.p. } 1 - \epsilon. \end{cases}$$

We use the same mixture weights as Sherlock et al. (2013) and set  $\epsilon$  to be 0.1. The

acceptance probability is given by:

$$\alpha\left(\phi_r^{(m-1)}, \phi_r^{(m)*}\right) = \min\left(1, \frac{p\left(\phi_r^{(m)*} | \mathbf{b}_c^{(m-1)}, \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q\right)}{p\left(\phi_r^{(m-1)} | \mathbf{b}_c^{(m-1)}, \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q\right)}\right).$$

The first component of the mixture is a multivariate normal with a fixed scaling factor and covariance matrix.  $\lambda^{(0)}$  is set to  $2.38/\sqrt{p_r + 2}$ , the theoretically optimal scaling factor for a target which is the product of identical normal densities (Roberts and Rosenthal, 2001). Under any reasonable prior for  $\phi_r$  and given a large amount of data,  $\text{MVN}(\hat{\phi}_r, \hat{\Sigma}_{\phi_r})$  provides a reasonable approximation to the posterior, where  $\hat{\Sigma}_{\phi_r}$  is the inverse of the Observed Fisher Information matrix from ML estimation. Hence  $\hat{\Sigma}_{\phi_r}$  is used for  $\Sigma^{(0)}$ . The second component is an adaptive proposal distribution, where the scaling factor and variance matrix are updated with each iteration of the chain, based on the previously sampled values (Atchadé et al., 2009):

$$\boldsymbol{\mu}^{(m)} = \frac{m-1}{m} \boldsymbol{\mu}^{(m-1)} + \frac{1}{m} \phi_r^{(m)},$$

$$\Sigma^{(m)} = \frac{m-1}{m} \Sigma^{(m-1)} + \frac{1}{m} \left( \phi_r^{(m)} - \boldsymbol{\mu}^{(m-1)} \right) \left( \phi_r^{(m)} - \boldsymbol{\mu}^{(m-1)} \right)',$$

$$\lambda^{(m)} = \begin{cases} \lambda^{(m-1)} + 2.3 \frac{\lambda^{(0)}}{100m^{0.85}} & \text{if } \phi_r^{(m)*} \text{ is accepted,} \\ \lambda^{(m-1)} - \frac{\lambda^{(0)}}{100m^{0.85}} & \text{if } \phi_r^{(m)*} \text{ is rejected.} \end{cases}$$

Proposals from the adaptive component of the mixture are only allowed after 1000 accepted jumps from the standard RWM, in order to avoid obtaining a singular variance matrix for  $\Sigma^{(m)}$ . As  $m \rightarrow \infty$ , changes to the scaling factor between iterations become smaller, leading to an equilibrium acceptance rate of 0.3. This is part of the diminishing adaptation condition, which requires vanishingly small changes to the transition kernel in order to preserve stationarity (Sherlock et al., 2010).

### 5.2.4 Chain diagnostics

The algorithm was coded in R and updates of the customer random effects were parallelised over four CPU cores. On a computer with an Intel Nehalem 1.60 GHz CPU, 100,000 iterations of the chain took approximately one week to run. Three independent Markov chains of length 100,000 were generated from different starting positions, sampled from the prior distribution. Each chain was thinned by keeping every 10<sup>th</sup> set of sampled values. Figure 5.2.2 shows six of the trace plots from the first chain, which are representative of the other random effects and parameters sampled by the chain. The adaptive proposal for the referral price parameters is quick to learn the shape of the posterior and appears to provide good mixing.

The Gelman-Rubin statistic (Gelman and Rubin, 1992) compares within-chain variance to between-chain variance across multiple MCMC chains to assess convergence to the stationary distribution. The statistic was calculated from the three chains for each of the 53 components of  $\phi_r$  and 9,834 customer random effects for price. Figure 5.2.3 shows the median, maximum and maximum 97.5<sup>th</sup> quantile of the combined statistics at each iteration of the chain. Within the first few thousand iterations, the



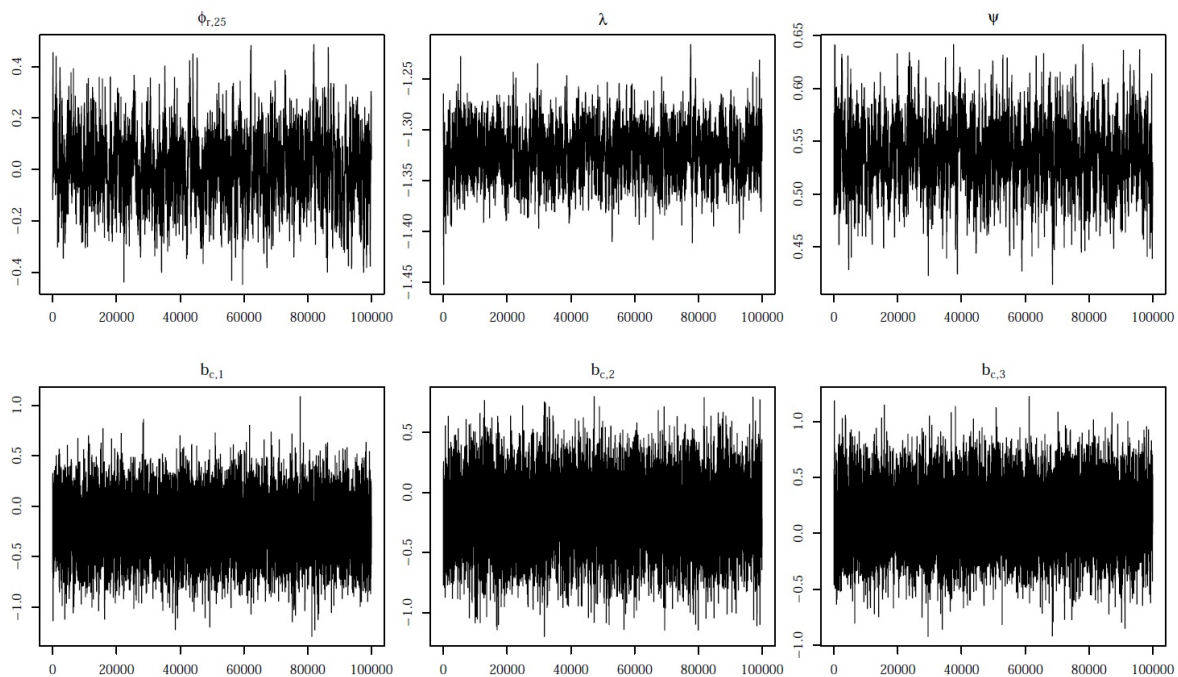


Figure 5.2.2: Example traceplots for the first of the three runs of the Markov chain. The top row shows three of the referral price parameters,  $\phi_r$ , and the bottom row shows three of the customer random effects,  $\mathbf{b}_c$ .

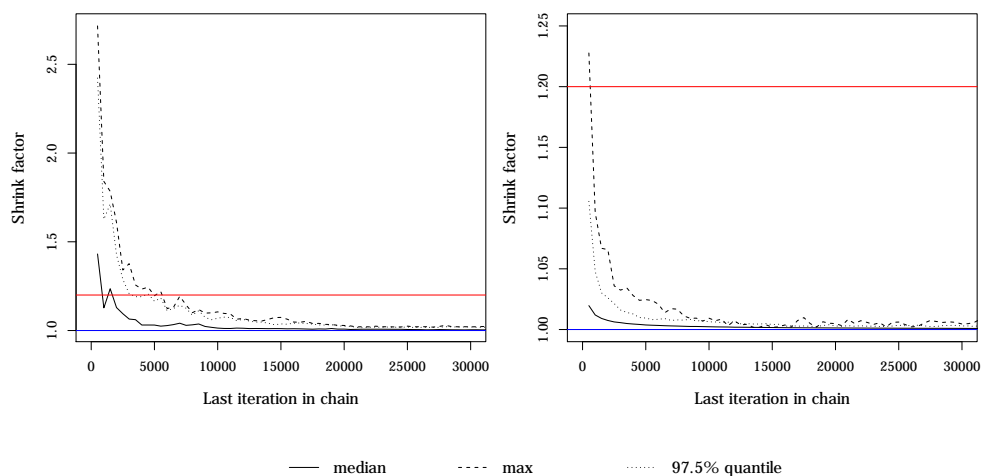


Figure 5.2.3: Combined Gelman-Rubin statistics for all 53 referral price parameters,  $\phi_r$  (left) and all 9,834 customer random effects,  $\mathbf{b}_c$ . The ideal statistic of 1 (—) and the threshold of 1.2 (—) are also shown.

statistics of the customer random effects fall below the threshold of 1.2 suggested in Gelman (1996). The plot for  $\phi_r$  suggests a burn-in period of 10,000 iterations should be sufficient. Inference is therefore performed using the final 90,000 iterations from each of the runs combined, which gives a sample of 27,000 since the chains have been thinned by a factor of 10.

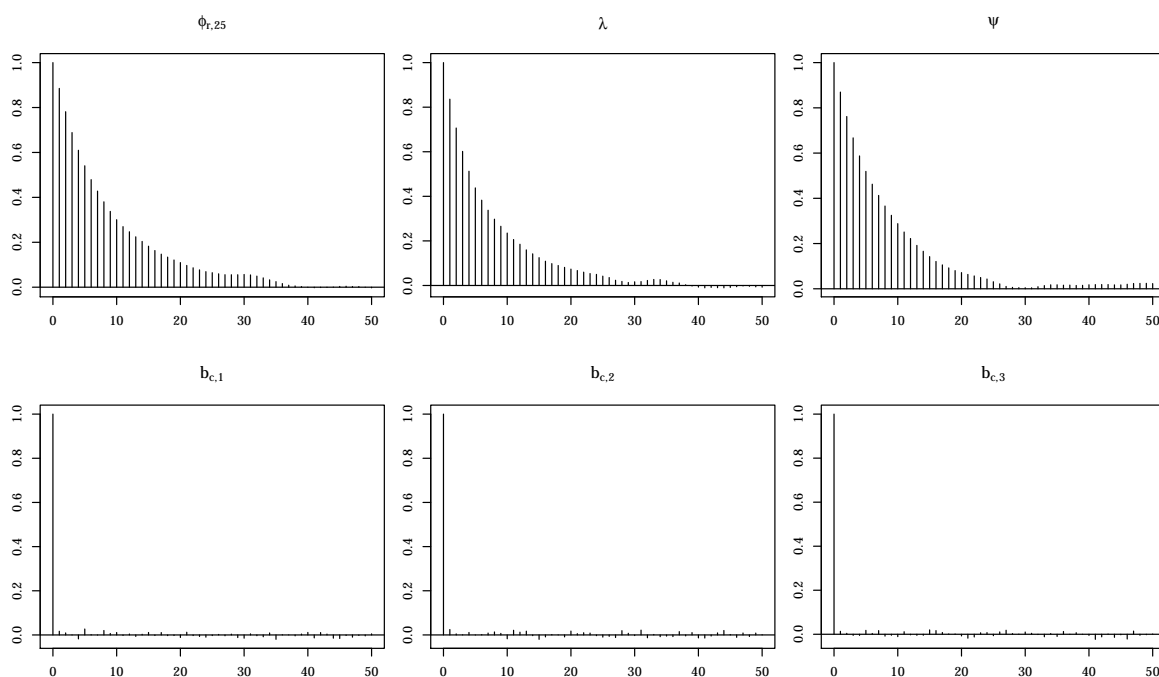


Figure 5.2.4: Example autocorrelation plots for the first of the three runs of the Markov chain. The top row shows three of the referral price parameters,  $\phi_r$ , and the bottom row shows three of the customer random effects,  $\mathbf{b}_c$ .

Figure 5.2.4 plots the autocorrelation function up to a lag of 50 for six of the components from the first MCMC run. The simulated values for each customer random effect resemble an independent sample, whereas the dimension of  $\phi_r$  leads to persistent autocorrelation in the chain. Even at a lag of 10, which corresponds to values separated by 100 iterations, autocorrelation is above 0.2. Table 5.2.2 shows a sum-

mary of the effective sample sizes (ESS) obtained by combining the three chains. The effective sample sizes for the referral price parameters are between 1,500 and 2,200, while the customer random effects all have an ESS above 20,000. For predictions of profit in Section 5.3, the posterior sample is thinned further to 3,000 in order to reduce the computational cost of marginalising over the parameters. The impact on effective sample size is minor.

Table 5.2.2: Effective sample sizes

		Sample size	Effective sample size		
			Minimum	Median	Maximum
Full sample	$\phi_r$	27,000	1,559	1,811	2,116
	$\mathbf{b}_c$	27,000	22,478	25,689	26,718
Reduced sample	$\phi_r$	3,000	1,423	1,695	1,909
	$\mathbf{b}_c$	3,000	2,532	3,000	3,000

### 5.2.5 Parameter estimates

Bayesian estimates of the referral price parameters and customer random effects for price are given by the sample posterior means:

$$\mathbb{E}[\phi_r | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q] = \frac{1}{27,000} \sum_{m=1}^{27,000} \phi_r^{(m)},$$

$$\mathbb{E}[b_{c,j} | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_q] = \frac{1}{27,000} \sum_{m=1}^{27,000} b_{c,j}^{(m)}.$$

Figure 5.2.5 shows kernel density plots for posterior samples of three referral price parameters and three customer random effects. The examples do not display any notable asymmetry, however the posterior mean for the price coefficient,  $\lambda$ , differs slightly from the MLE (provided in Section 4.4.4). This was true for several of the

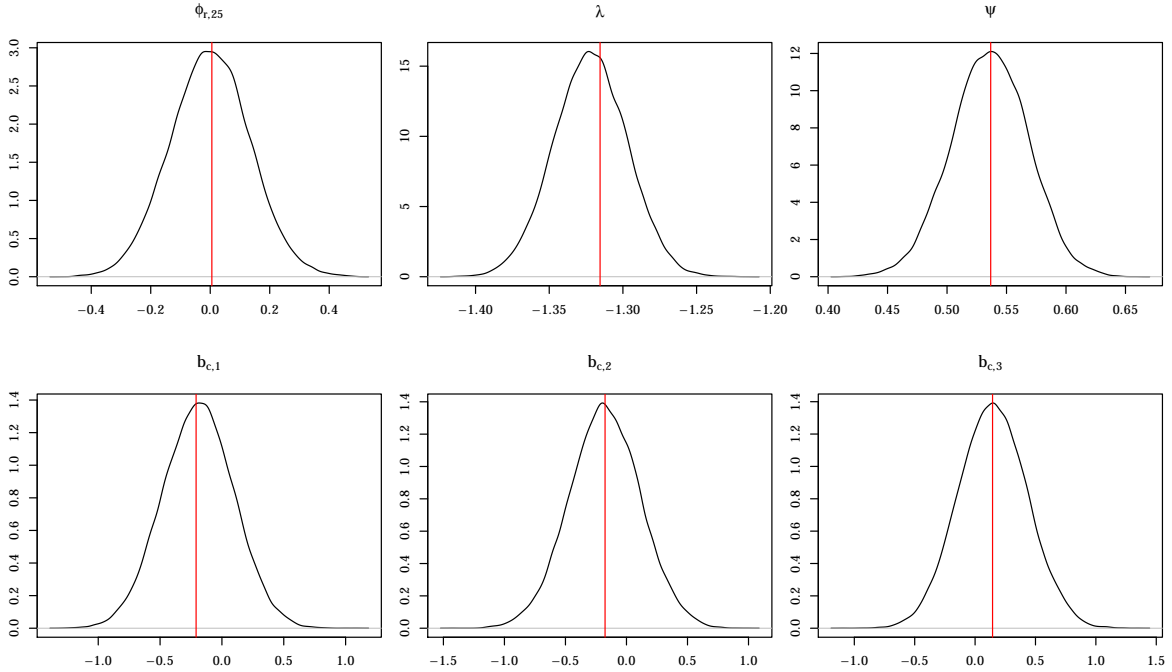


Figure 5.2.5: Example kernel density plots for the posterior samples of parameters. The top row shows three of the referral price parameters,  $\phi_r$ , and the bottom row shows three of the customer random effects,  $\mathbf{b}_c$ . Corresponding MLEs and BLUPs are shown by a vertical line (—)

parameters in the sample.

Figure 5.2.6 shows the best linear unbiased predictors for the customer random effects and corresponding posterior means. The BLUPs are the prior means of the customer effects, given by the probability distribution for  $\mathbf{b}_c$ , conditional on the observed quote prices. The posterior is produced by updating the prior to include information from customer responses, and the plot shows little change in the estimates of the random effects. However, whilst the prediction intervals for the BLUPs have a mean width of 0.93, the mean width of the highest posterior density (HPD) interval is 1.13. Of the customer effects with a relatively wider HPD interval, the majority are those with a large absolute estimate,  $|\hat{b}_{c,j}| > 0.8$ . This suggests that the piecewise linear

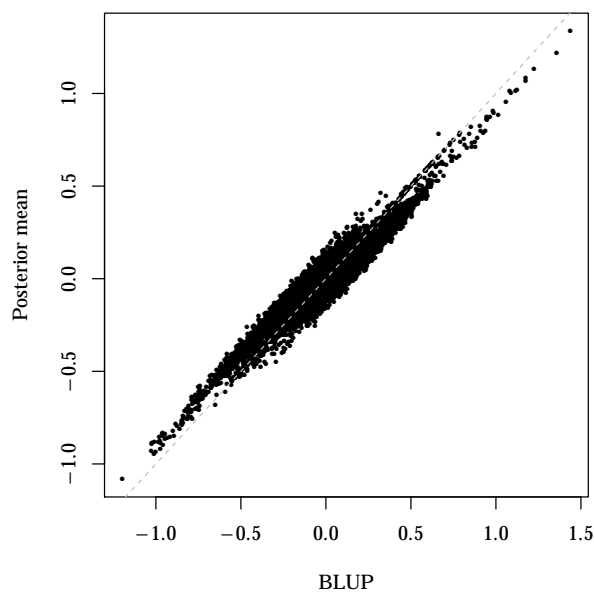


Figure 5.2.6: Best linear unbiased predictors of the customer random effects for price,  $\hat{\mathbf{b}}_c$ , plotted against corresponding posterior means, with a (0,1) line for reference (-----).

transformation in the referral price model is pulling the customer random effects in a different direction to that of the quote price model.

Maximum likelihood estimates and sample posterior means of the raw referral price parameters are plotted in Figure 5.2.7. The estimates are mostly in agreement, with an average difference of 0.008, and all of the 95% credible intervals include the maximum likelihood estimate of the parameter. The largest differences between the MLEs and Bayesian estimates occur for `quoteHist`, the quote price history coefficients, with a maximum difference of 0.073. Given that the smallest categories for this factor concerns just 122 quotes, we can expect larger discrepancies here. The confidence intervals and HPD intervals are similar in size, and parameters with confidence intervals containing zero also have HPD intervals containing zero.

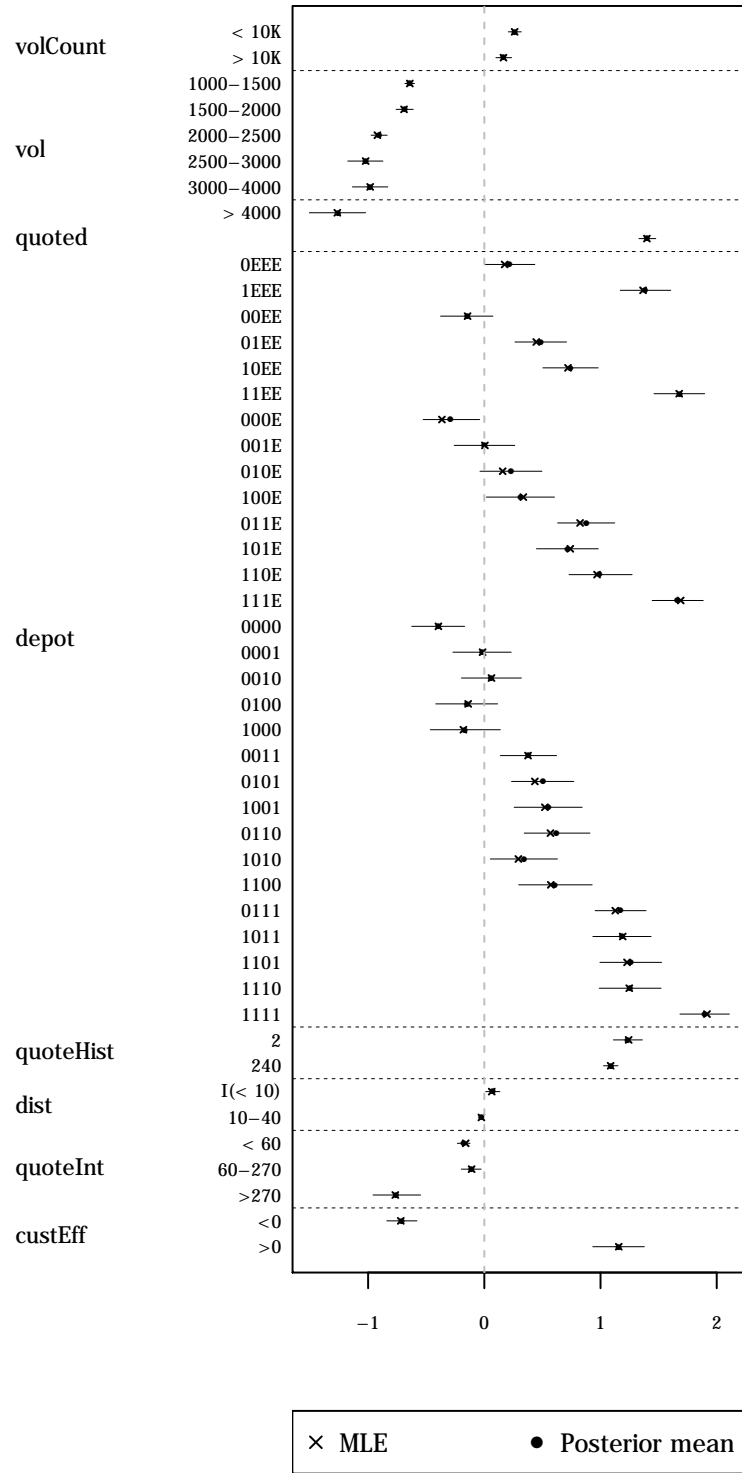


Figure 5.2.7: Posterior means and 95% HPD intervals for the raw referral price parameters,  $\phi_r$ . ML estimates are shown for comparison.

### 5.3 Counterfactual analysis

Using the posterior sample of parameters, we now investigate how changes to the price of previous quotes would impact profit. It is important to note here that we restrict our focus to the most recent quote recorded for each customer in Subset 3. The reason for this is that the customer's previous responses to quotes are included as a categorical covariate in the model, `quoteHist`. Hence a change in the price quoted to a customer in an earlier quote may shift their response, and thus change the customer's quote history for later quotes, which would in turn affect predictions of willingness to pay. It is possible to sequentially modify the quotes for each customer, by simulating a response to their first quote and then updating their quote history, before modifying and simulating responses to subsequent quotes. However, since the referral price coefficients have been estimated from customer responses to the observed prices, it was thought to be more prudent to avoid modifying covariates and focus on each customer's last quote. This leaves a sample of  $\tilde{N} = 9,834$  quotes for analysis.

#### 5.3.1 Expected profit as a model check

Section 5.1.2 defined profit improvement,  $\Delta W(\tilde{\mathbf{q}}, \mathbf{q})$ , as the difference between profit under a new set of prices,  $W(\tilde{\mathbf{q}})$  and actual profit at original prices,  $W(\mathbf{q})$ . As a model check we can also compare expected profit at original prices,  $\mathbb{E}_{\mathbf{Y}|\mathbf{q}}[W(\mathbf{q})]$ , with actual profit.

$$\mathbb{E}_{\mathbf{Y}|\mathbf{q}}[W(\mathbf{q})] = \sum_{i=1}^{\tilde{N}} h(q_i) \cdot \pi(q_i; \dot{\mathbf{x}}_{r,i}, \mathbf{b}_c, \boldsymbol{\phi}_r).$$

The expectation is taken with respect to  $\mathbf{Y}$ , the customer responses to  $\mathbf{q}$ . Comparison to actual profit then indicates whether predictions of willingness to pay are in line with observed customer responses. Expected profit is estimated as follows:

$$\hat{\mathbb{E}}_{\mathbf{Y}|\mathbf{q}}[W(\mathbf{q})] = \sum_{i=1}^{\tilde{N}} h(q_i) \cdot \hat{\pi}(q_i; \dot{\mathbf{x}}_{r,i}).$$

where,

$$\begin{aligned} \hat{\pi}(q_i; \dot{\mathbf{x}}_{r,i}) &= \hat{\mathbb{E}}_{\mathbf{b}_c, \phi_r | \mathbf{y}, \mathbf{q}}[\pi(q_i; \dot{\mathbf{x}}_{r,i}, \mathbf{b}_c, \phi_r)] \\ &= \frac{1}{M} \sum_{m=1}^M \pi(q_i; \dot{\mathbf{x}}_{r,i}, \mathbf{b}_c^{(m)}, \phi_r^{(m)}). \end{aligned} \quad (5.3.1)$$

The term (5.3.1) provides a prediction of willingness to pay for each quote, which marginalises over the parameters. Willingness to pay is calculated for each set of sampled values for the referral price parameters and random effects from the reduced Markov chain, and then averaged. The variance of the profit under original prices is estimated in a similar fashion,

$$\widehat{\text{Var}}_{\mathbf{Y}|\mathbf{q}}[W(\mathbf{q})] = \sum_{i=1}^{\tilde{N}} h(q_i)^2 \cdot \widehat{\text{Var}}(\pi(q_i; \dot{\mathbf{x}}_{r,i})),$$

where

$$\widehat{\text{Var}}(\pi(q_i; \dot{\mathbf{x}}_{r,i})) = \frac{1}{M} \sum_{m=1}^M \pi(q_i; \dot{\mathbf{x}}_{r,i}, \mathbf{b}_c^{(m)}, \phi_r^{(m)}) \left(1 - \pi(q_i; \dot{\mathbf{x}}_{r,i}, \mathbf{b}_c^{(m)}, \phi_r^{(m)})\right).$$

The estimates in Table 5.3.1 show that observed profit is well within two standard



deviations of expected profit.

Table 5.3.1: Expected profit under original prices

$W(\mathbf{q})$	$\hat{\mathbb{E}}_{\mathbf{Y} \mathbf{q}}[W(\mathbf{q})]$	$\sqrt{\widehat{\text{Var}}_{\mathbf{Y} \mathbf{q}}[W(\mathbf{q})]}$	95% CI for $\mathbb{E}_{\mathbf{Y} \mathbf{q}}[W(\mathbf{q})]$
£ 147,387	£ 150,769	£ 2,711	(£145,456, £156,082)

### 5.3.2 Predicting profit improvement

We now outline a procedure for predicting profit under a new set of prices,  $W(\tilde{\mathbf{q}})$ .

Willingness to pay under an alternative price (5.1.10) was defined in Section 5.1.1

and for current purposes, it is helpful to reparameterise the function in terms of the

untransformed referral price parameters,  $\phi_r$ :

$$\tilde{\pi}(\tilde{q}; y, q, \dot{\mathbf{x}}_r, \mathbf{b}_c, \phi_r) = \begin{cases} \frac{\text{P}(R > \tilde{q} | \mathbf{x}_r)}{\text{P}(R > q | \mathbf{x}_r)} & \text{if } \tilde{q} > q, y = 1, \\ 0 & \text{if } \tilde{q} > q, y = 0, \\ 1 & \text{if } \tilde{q} < q, y = 1, \\ \frac{\text{P}(R \leq q | \mathbf{x}_r) - \text{P}(R \leq \tilde{q} | \mathbf{x}_r)}{\text{P}(R \leq q | \mathbf{x}_r)} & \text{if } \tilde{q} < q, y = 0, \end{cases} \quad (5.3.2)$$

where,

$$\text{P}(R > \tilde{q} | q, \mathbf{x}_r) = \Phi(\mathbf{x}_r' \gamma - \lambda \tilde{q} + \psi \hat{\varepsilon}_q),$$

$$\text{P}(R > \tilde{q} | q, \mathbf{x}_r) = \Phi(-[\mathbf{x}_r' \gamma - \lambda \tilde{q} + \psi \hat{\varepsilon}_q]).$$

Here  $\mathbf{x}'_r \gamma$  can be written as  $\dot{\mathbf{x}}'_r \gamma_1 + [k(\mathbf{z}'_c \mathbf{b}_c)]' \gamma_2$  in the same way as (5.2.4). In order to simplify the expressions below,  $\tilde{\pi}(\tilde{q}_i; y_i, q_i, \dot{\mathbf{x}}_{r,i}, \mathbf{b}_c^{(m)}, \phi_r^{(m)})$  is denoted by  $\tilde{\pi}_i^{(m)}$ . This represents willingness to pay at the new price for quote  $i$ , conditional on a set of parameter values,  $\{\mathbf{b}_c^{(m)}, \phi_r^{(m)}\}$ . The posterior sample of referral price parameters and customer price effects is used to perform inference on the posterior predictive distribution for customer responses and profit, as described below.

### Predicting profit

For each set of sampled parameter values from the posterior,  $m = 1, \dots, M$ ,

For each quote,  $i = 1, \dots, \tilde{N}$ ,

Calculate willingness to pay at  $\tilde{q}_i, \tilde{\pi}_i^{(m)}$ .

Sample  $\tilde{Y}_i$  from the posterior predictive distribution:

$$\tilde{y}_i^{(m)} | \tilde{q}_i, y_i, q_i \sim \text{Bernoulli}(\tilde{\pi}_i^{(m)}).$$

Sum over quotes,

$$W^{(m)}(\tilde{\mathbf{q}}) = \sum_{i=1}^{\tilde{N}} h(\tilde{q}_i) \cdot \tilde{y}_i^{(m)}.$$

Average over the posterior sample to estimate total profit at new prices:

$$\hat{\mathbb{E}}[W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q}] = \frac{1}{M} \sum_{m=1}^M W^{(m)}(\tilde{\mathbf{q}}). \quad (5.3.3)$$

Here,  $\{W^{(m)}(\tilde{\mathbf{q}})\}_{m=1}^M$  is a sample from the posterior predictive distribution for profit under new prices, and (5.3.3) is a Monte Carlo estimator of the expected profit with respect to this posterior predictive. Then profit improvement, or the impact on profit as a result of changing prices from  $\mathbf{q}$  to  $\tilde{\mathbf{q}}$ , is estimated by:

$$\hat{\mathbb{E}}[\Delta W(\tilde{\mathbf{q}}, \mathbf{q})] = \hat{\mathbb{E}}[W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q}] - W(\mathbf{q}). \quad (5.3.4)$$

Our uncertainty in the improvement to profit is also of interest, since a high posterior probability of negative profit improvement is indicative of a poor pricing policy. A Monte Carlo estimate of the variance in profit improvement is given by the variance of the posterior predictive sample for profit:

$$\begin{aligned} \widehat{\text{Var}}[\Delta W(\tilde{\mathbf{q}}, \mathbf{q})] &= \widehat{\text{Var}}[W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q}], \\ &= \frac{1}{M} \sum_{m=1}^M \left( W^{(m)}(\tilde{\mathbf{q}}) - \hat{\mathbb{E}}[W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q}] \right)^2 \end{aligned} \quad (5.3.5)$$

A credible interval is used to assess whether there is a significant probability that a price change will negatively impact profit,  $\Delta W(\tilde{\mathbf{q}}, \mathbf{q}) < 0$ . Actual profit is subtracted from  $\{W^{(m)}(\tilde{\mathbf{q}})\}_{m=1}^M$  and the quantiles of this sample are used to calculate credible intervals for profit improvement under new prices.

By the law of total variance (e.g. Weiss et al. (2005)), the variance of profit under new prices can be decomposed as follows:

$$\text{Var}(W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q}) = \mathbb{E}[\text{Var}(W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q}, \boldsymbol{\phi}_r, \mathbf{b}_c)] + \text{Var}(\mathbb{E}[W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q}, \boldsymbol{\phi}_r, \mathbf{b}_c]) \quad (5.3.6)$$

The first term on the right gives the expected variance of profit and captures variability from the customer responses, after averaging over parameters. The second term arises from variability in the expected values of each customer response, and hence represents parameter uncertainty. Expected profit under new prices, conditional on the parameters, is given by:

$$\mathbb{E} [W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q}, \boldsymbol{\phi}_r, \mathbf{b}_c] = \sum_{i=1}^{\tilde{N}} h(\tilde{q}_i) \tilde{\pi}(\tilde{q}_i; y_i, q_i, \dot{\mathbf{x}}_{r,i}, \mathbf{b}_c, \boldsymbol{\phi}_r).$$

And the variance of profit under new prices, conditional on the parameters, is given by:

$$\text{Var} (W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q}, \boldsymbol{\phi}_r, \mathbf{b}_c) = \sum_{i=1}^{\tilde{N}} h(\tilde{q}_i)^2 \tilde{\pi}(\tilde{q}_i; y_i, q_i, \dot{\mathbf{x}}_{r,i}, \mathbf{b}_c, \boldsymbol{\phi}_r) (1 - \tilde{\pi}(\tilde{q}_i; y_i, q_i, \dot{\mathbf{x}}_{r,i}, \mathbf{b}_c, \boldsymbol{\phi}_r)).$$

Both of these terms are estimated by Monte Carlo estimation, using the posterior sample of parameters:

$$\hat{\mathbb{E}} [\text{Var} (W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q}, \boldsymbol{\phi}_r, \mathbf{b}_c)] = \frac{1}{M} \sum_{m=1}^M \sum_{i=1}^{\tilde{N}} h(\tilde{q}_i)^2 \tilde{\pi}_i^{(m)} (1 - \tilde{\pi}_i^{(m)}) \quad (5.3.7)$$

$$\widehat{\text{Var}} (\mathbb{E} [W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q}, \boldsymbol{\phi}_r, \mathbf{b}_c]) = \frac{1}{M} \sum_{m=1}^M \left( \sum_{i=1}^{\tilde{N}} h(\tilde{q}_i) \tilde{\pi}_i^{(m)} - \left[ \frac{1}{M} \sum_{m=1}^M \sum_{i=1}^{\tilde{N}} h(\tilde{q}_i) \tilde{\pi}_i^{(m)} \right] \right)^2 \quad (5.3.8)$$

The relative sizes of these two terms indicate which of the unknowns - customer responses or parameters - has a larger impact on variability in profit. The sum of (5.3.7) and (5.3.8) gives an estimate of the total variance in profit improvement,

$\widehat{\text{Var}}^{\text{tot}}(W(\tilde{\mathbf{q}}) | \mathbf{y}, \mathbf{q})$ , and should equate to the variance calculated directly from the posterior sample for profit, (5.3.5).

The next section describes a set of price changes and for each of these, profit improvement is predicted in the manner outlined here. The results are presented in Section 5.3.4 and discussed in the following section.

### 5.3.3 Alternative pricing policies

We apply a number of price changes to the quotes in the sample, based on inferred purchase behaviour from the fitted model. The retail prices of either the full set or a subset of quotes are adjusted by a relative amount,  $(1 + \delta)$ , and then transformed according to (5.1.11) to give a new set of prices,  $\tilde{\mathbf{q}}$ . Only small price changes,  $\delta = \pm 0.01$ , are considered, in order to ensure that predictions are valid. The details of each price change are described in Table 5.3.2. As well as universal price changes, we test depot-specific price changes, where the quotes for each of the depots in Subset 3 are modified in isolation.

Customer-related covariates were used to specify price changes and in particular, we consider covariate effects which were found to be significant in the referral price model (5.2.2), but not in the quote price model (5.2.1). The reason for this, is that such covariates were shown to affect willingness to pay and may have been overlooked by sales agents in their pricing decisions. One example of this is the quote history of each customer, `quoteHist`. In Section 4.4.1 it was noted that the timing of previous acceptances has a strong effect on referral price compared to quote price. For instance,

in the 100E and 001E categories, both customers have three previous quotes, of which one was accepted. However in the first category, customers accepted their most recent quote and are more likely to accept their next quote, whereas customers in the second category accepted their oldest quote and are less likely to accept their quote. There may be an opportunity for sales agents to vary prices to match this behaviour and improve profit.

The number of days since the customer's previous quote, `quoteInt`, was also found to affect referral price but not quote price. And hence prices are modified according to the estimated relationship between `quoteInt` and willingness to pay. Quotes requested by customers with a more recent previous quote are given higher prices, whilst less active customers are offered lower prices. Similarly, the relationship between the customer's random effect for price and willingness to pay is used to determine suitable price changes. Customers with a large negative random effect,  $b_C \ll 0$ , are generally offered prices below average, however Figure 4.4.3 shows that the same customers are more likely to accept their quotes. The same is true for customers with a large positive price effect,  $b_c \gg 0$ . Hence prices for these customers are increased, whilst prices for customers with average-priced quotes are decreased.

The final policy is designed to compare differential price increases against a uniform price increase. Under a differential price increase, prices for some fraction of the quotes,  $f_1$ , are increased by  $\delta_1\%$  and prices for another fraction of quotes,  $f_2$ , are increased by  $\delta_2\%$ . This is compared to a uniform price increase for both sets of quotes by  $\kappa\%$ , where  $\kappa$  is a weighted average of the individual increases:

Table 5.3.2: Details of pricing policies

<b>Universal</b>
A 1% increase and decrease in the prices of all quotes.
<b>Depot</b>
A 1% increase and decrease in the prices of quotes for each depot, in turn.
<b>Quote history</b>
Price changes according to <code>quoteHist</code> , the customer's previous 4 responses to quotes
<ul style="list-style-type: none"> <li>a Customers with 2 previous quotes and one acceptance: decrease quotes for '01EE' by 1% and increase quotes for '10EE' by 1%.</li> <li>b Customers with 3 previous quotes and one acceptance: decrease quotes for '001E' by 1% and increase quotes for '100E' by 1%.</li> <li>c Customers with 3 previous quotes and 2 acceptances: decrease quotes for '011E' by 1% and increase quotes for '110E' by 1%.</li> <li>d Customers with 4 previous quotes and one acceptance: decrease quotes for '0001' by 1% and increase quotes for '1000' by 1%.</li> <li>e Customers with 4 previous quotes and 2 acceptances: decrease quotes for '0011' by 1% and increase quotes for '1100' by 1%.</li> <li>f Customers with 4 previous quotes and 3 acceptances: decrease quotes for '0111' by 1% and increase quotes for '1110' by 1%.</li> </ul>
<b>Quote interval</b>
Price changes according to <code>quoteInt</code> , the no. of days since the customer's last quote
<ul style="list-style-type: none"> <li>a Customers who request quotes relatively frequently: increase quotes where <math>\text{quoteInt} \leq 60</math> by 1%.</li> <li>b Customers who request quotes relatively infrequently: decrease quotes where <math>\text{quoteInt} &gt; 270</math> by 1%.</li> </ul>
<b>Customer random effect for price</b>
Price changes according to $b_c$
<ul style="list-style-type: none"> <li>a Customers with below-average prices: increase quotes where <math>b_c \leq -0.4</math> by 1%.</li> <li>b Customers with above-average prices: increase quotes where <math>b_c &gt; 0.2</math> by 1%.</li> <li>c Customers with average prices: decrease quotes where <math>-0.4 &lt; b_c \leq 0.2</math> by 1%.</li> </ul>
<b>Differential versus uniform price increases</b>
Individual prices increases and averaged price increases for Depots 240 and 260
<ul style="list-style-type: none"> <li>a Increase quotes for Depot 240 by 2% and quotes for Depot 260 by 1%.</li> <li>b Increase quotes for either Depot 240 or 260 by <math>\kappa\%</math> where <math>\kappa</math> is a weighted average of the individual differences applied in a.</li> </ul>

$$\kappa = \frac{f_1\delta_1 + f_2\delta_2}{f_1 + f_2}.$$

These price changes are applied to the quotes from the two largest depots, 240 and 260. Prices for the largest depot, 260, are increased by 1% whilst prices for depot 240 are increased by 2%, and then prices for both depots are increased by a weighted average, 1.2%.

### 5.3.4 Results

The tables below give estimates of the posterior mean and median, and 95% HPD intervals for profit improvement under each price change scenario. For the universal price changes, the variance of profit under the new prices was decomposed according to (5.3.6) and the associated estimates are given in Table 5.3.4, alongside the estimated posterior variance. Kernel density plots of the posterior sample for  $\Delta W(\mathbf{q}, \mathbf{q})$  under two of the alternative pricing policies are shown in Figures 5.3.1 and 5.3.2.

Table 5.3.3: Posterior estimates of profit improvement (£) after universal price changes

$\Delta W(\tilde{\mathbf{q}}, \mathbf{q})$	1% price increase	1% price decrease
Mean	7,860	-9,133
Med.	7,884	-9,133
HPD Int.	(6,270, 9,362)	(-10,449, -7,838)



Table 5.3.4: Variance decomposition for profit improvement (£) after universal price changes

	1% price increase	1% price decrease
$\hat{\mathbb{E}} [\text{Var} (W(\tilde{\mathbf{q}})   \mathbf{y}, \mathbf{q}, \phi_r, \mathbf{b}_c)]$	555,465	442,463
$\widehat{\text{Var}} (\mathbb{E} [W(\tilde{\mathbf{q}})   \mathbf{y}, \mathbf{q}, \phi_r, \mathbf{b}_c])$	53,546	28,093
$\widehat{\text{Var}}^{\text{tot}} (W(\tilde{\mathbf{q}})   \mathbf{y}, \mathbf{q})$	609,010	471,556
$\widehat{\text{SD}}^{\text{tot}} (W(\tilde{\mathbf{q}})   \mathbf{y}, \mathbf{q})$	780	687
$\widehat{\text{Var}} (W(\tilde{\mathbf{q}})   \mathbf{y}, \mathbf{q})$	629,979	457,084
$\widehat{\text{SD}} (W(\tilde{\mathbf{q}})   \mathbf{y}, \mathbf{q})$	794	676

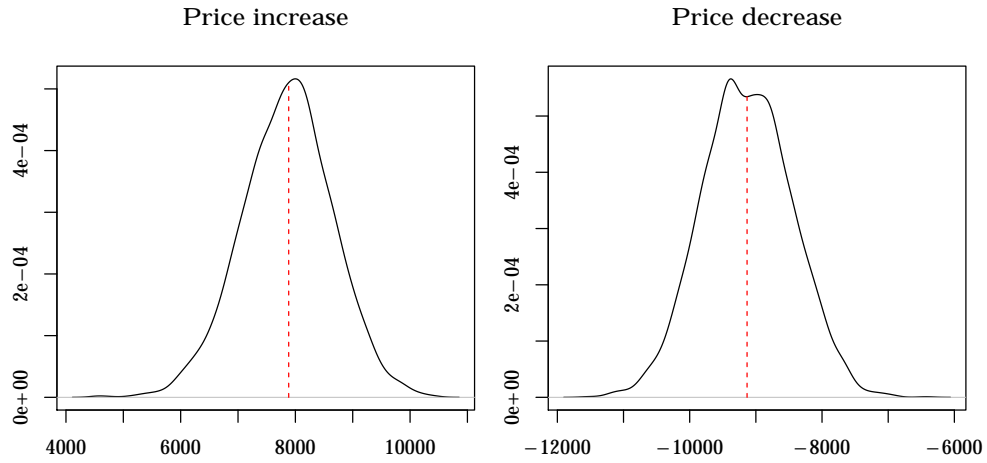


Figure 5.3.1: Kernel density plots for posterior samples of profit improvement after a 1% increase in retail prices (left) and a 1% decrease in prices (right).

Table 5.3.5: Posterior estimates of profit improvement (£) after price changes for each depot

Depot	$\Delta W(\tilde{\mathbf{q}}, \mathbf{q})$	1% price increase	1% price decrease
2	Mean	496	-634
	Med.	516	-650
	HPD Int.	(67, 901)	(-917, -291)
240	Mean	2,267	-2,582
	Med.	2,290	-2,590
	HPD Int.	(1,443, 3,066)	(-3,251, -1,847)
260	Mean	5,070	-5,878
	Med.	5,097	-5,896
	HPD Int.	(3,869, 6,150)	(-6,961, -4,806)

Table 5.3.6: Posterior estimates of profit improvement (£) after differentiated and uniform price changes for Depots 240 and 260

$\Delta W(\tilde{\mathbf{q}}, \mathbf{q})$	Differential price increase	Uniform price increase
Mean	9,117	8,468
Med.	9,122	8,495
HPD Int.	(7,414, 10,848)	(6,820, 10,044)

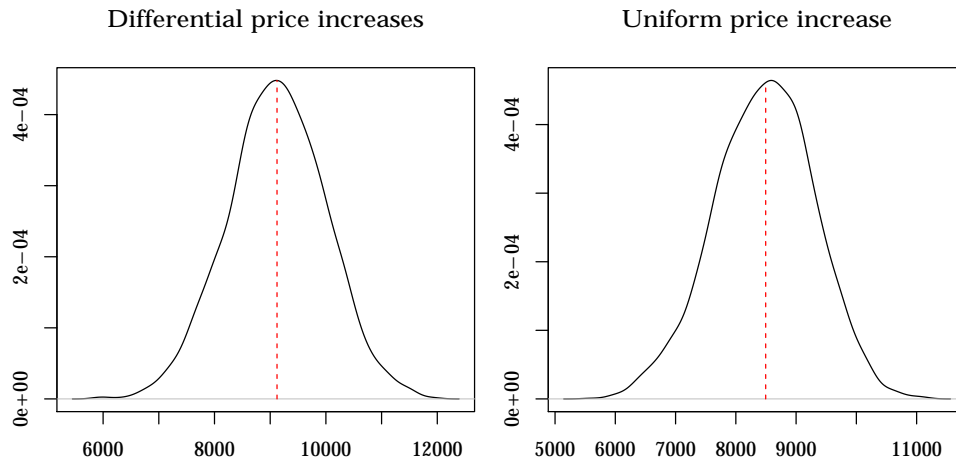


Figure 5.3.2: Kernel density plots for posterior samples of profit improvement after applying differentiated (left) and uniform (right) price changes to Depots 240 and 260.

Table 5.3.7: Posterior estimates of profit improvement (£) after price changes based on `quoteInt`

$\Delta W(\tilde{\mathbf{q}}, \mathbf{q})$	1% price increase for <code>qint</code> $\leq 60$	1% price decrease for <code>qint</code> $> 270$
Mean	1,326	-1,323
Med.	1,352	-1,334
HPD Int.	(678, 1,918)	(-1,736, -898)

Table 5.3.8: Posterior estimates of profit improvement (£) after price changes based on `quoteHist`

$\Delta W(\tilde{\mathbf{q}}, \mathbf{q})$	a. 2 previous quotes, 1 acceptance	b. 3 previous quotes, 1 acceptance	c. 3 previous quotes, 2 acceptances
Mean	-111	-23	-221
Med.	-113	-32	-226
HPD Int.	(-414, -182)	(-119, 99)	(-337, -24)
$\Delta W(\tilde{\mathbf{q}}, \mathbf{q})$	d. 4 previous quotes, 1 acceptance	e. 4 previous quotes, 2 acceptances	f. 4 previous quotes, 3 acceptances
Mean	-221	-178	-840
Med.	-226	-177	-853
HPD Int.	(-336, -24)	(-356, 6)	(-1,136, -523)

Table 5.3.9: Posterior estimates of profit improvement (£) after price changes based on  $\mathbf{b}_c$ 

$\Delta W(\tilde{\mathbf{q}}, \mathbf{q})$	1% price increase for $b_c \leq -0.4$	1% price increase for $b_c > 0.2$	1% price decrease for $-0.4 < b_c \leq 0.2$
Mean	269	-1,187	6,782
Med.	281	-1,166	6,771
HPD Int.	(-169, 639)	(-1,990, -393)	(5,773, 7,767)

### 5.3.5 Discussion of results

Table 5.3.3 presents the posterior mean and median, as well as the 95% HPD interval for profit improvement after a universal price increase and decrease. Figure 5.3.1 shows kernel density plots of the posterior samples for profit improvement in each case. The

plots display some asymmetry, which suggests that prediction intervals based on a normal approximation would be unsuitable here. A universal price increase of 1% is estimated to increase profit by around £7,800, an average of £0.80 per quote, with an minimum predicted increase of £6,270. In contrast, a price decrease leads to an estimated loss of £9,100. This supports the empirical evidence of sales agents offering prices below their profit-maximising levels (Garrow et al., 2006).

For the universal price changes, variance in predicted profit was decomposed into variability arising from parameter uncertainty,  $\mathbb{E}(\text{Var}(W(\tilde{\mathbf{q}})|\mathbf{y}, \mathbf{q}, \phi_r, \mathbf{b}_c))$ , and variability caused by randomness in customer responses,  $\text{Var}(\mathbb{E}[W(\tilde{\mathbf{q}})|\mathbf{y}, \mathbf{q}, \phi_r, \mathbf{b}_c])$ . The results are given in Table 5.3.4, and show that in both cases, variance is dominated by the latter term. The contribution from randomness in customer responses exceeds the contribution from parameter uncertainty by a factor of 10 under a price increase, and by a factor of 15 under a price decrease. This implies that the parameters are relatively well-identified, and expected profit,  $h(q) \pi(q; \mathbf{x})$ , can be estimated with high precision. The sum of the two variance components gives an estimate of the total variance in predicted profit,  $\widehat{Var}^{\text{tot}}$ . This is compared to the variance in the posterior sample for profit, generated by simulating customer responses from the posterior predictive distribution. The posterior variance estimate is shown in the last row of Table 5.3.4 and is close to  $\widehat{Var}^{\text{tot}}$ , as we would expect.

Variability in the volume of the order,  $V$ , also contributes to variance in predicted profit. For quotes which were accepted at original prices, the standard deviation of the volume requested is 569 litres, compared to a standard deviation of 488 litres for rejected quotes. Under a price increase, previously rejected quotes are rejected with

certainty, whereas responses to previously accepted quotes are simulated with a new willingness to pay. These quotes generally have a more variable order size and so the profit from these quotes is also more variable. As a result, we observe a relatively higher variance in predicted profit under a price increase, compared to a price decrease.

In Table 5.3.5, price changes have been applied to each depot in turn, leaving quotes from other depots unchanged. The estimated impact on profit follows the same direction as estimates for universal price changes, although the size of the impact differs between the depots. Over 75% of quotes in the sample are offered by Depot 260, however after prices are raised the average increase in profit per quote is £0.68, the lowest of the three depots. The medium-sized Depot 240 produces an increase of £1.30 per quote, whilst the smallest, Depot 2 produces an increase of £0.99.

Since the average increase in profit per quote differs by depot, this suggests that price rises should be tailored to each depot in order to maximise profit. The results in Table 5.3.6 confirm this idea. Here, differential price increases and uniform price increases were applied to the two largest depots, 240 and 260. In the first case, prices for Depot 260 were increased by 1% and prices for Depot 240 were increased by 2%. In the second case, prices for both depots were increased by a weighted average, 1.2%, where the weights were based on the fraction of quotes belonging to each depot. Figure 5.3.2 shows kernel density plots of the posterior sample for profit improvement in each case. The differential increase is estimated to have a larger impact on profit, however there is a large overlap between the prediction intervals.

Profit results for price changes based on the time interval between quotes are given in Table 5.3.7. The estimates suggest that a price increase for customers who

request quotes relatively frequently ( $\text{quoteInt} \leq 60$ ) would improve profit by £1,300. Whereas a price decrease for customers who request quotes relatively infrequently would decrease profit by the same amount. The price changes based on purchase history are all estimated to produce a negative impact on profit (Table 5.3.8) and with the exception of changes (b.) and (e.), the HPD intervals exclude zero. This is because each price change involves a mixture of price increases for some subset of quotes and price decreases for another subset of quotes. As we have established earlier, agents generally under-price quotes and so decreasing prices tends to negatively impact profit. The results suggest that agents are right in not lowering prices for customers who have recently rejected a quote, compared to customers who rejected older quotes.

Finally, Table 5.3.9 shows the estimated change in profit when prices are modified according to the customer random effect for price. Raising prices for customers who are generally offered prices below average ( $b_c \leq -0.4$ ) produces a minor improvement in profit. Since the prediction interval contains zero, this implies that customers in this group are price-sensitive and the seller may lose business from these customers if prices are raised. When prices are increased for customers who generally pay higher prices ( $b_c > 0.2$ ), profit is estimated to fall by £1,200. Hence we can conclude that price-insensitive customers are already being offered near-optimal prices. On the other hand, a price decrease for customers who are typically offered average prices ( $-0.4 < b_c \leq 0.2$ ) produces a substantial improvement in profit. This suggests that agents should lower prices for newer customers, who tend to fall in this group.

## 5.4 Summary

The model developed in Chapter 4 for the sale of heating oil was used to here to investigate the impacts of price changes on profit. Given the original price and customer response for each quote, we can define willingness to pay at an alternative price and simulate the customer's response to this new price. Since the parameters of the WTP function are not known with certainty, Bayesian estimation of the willingness to pay model was performed by adaptive MCMC in order to obtain a posterior sample of parameters. This sample was then used to simulate customer responses under a new price, and in doing so, average over any parameter uncertainty. From these customer responses we can predict profit at the new set of prices, and compare this to profit at the original prices in order to measure profit improvement.

In general, the results showed that agents tend to offer prices below their profit-maximising levels, since a 1% increase in price led to an estimated increase in profit of almost £8,000. Furthermore, there was evidence that differential price increases, whereby price increases are varied over different groups of quotes, leads to greater profit improvement than a uniform price increase. This indicates that the seller could benefit from targeted price increases, for example based on the number of days since the customer's previous quote or the depot offering the quote.

The analysis here treated the quote price model as fixed, and estimates of profit and willingness to pay were made conditionally on MLEs of the quote price parameters and BLUPs of the agent and day random effects, which are relatively well-identified. While the customer price effects and referral price parameters are known with less

precision, it was shown that relative to variability in customer responses, uncertainty in these parameters provides a smaller contribution to the variance of predictions. Nonetheless, a natural extension would be to perform MCMC estimation of the full model before simulating customer responses. The primary advantage of this approach would be that the quote price residuals, which enter into the WTP function to control for endogeneity, are estimated alongside the model parameters. Likewise, estimation of the customer price effect would use information from both the price offered and the customer response.



# Chapter 6

## Conclusions

In the concluding chapter, the main results of this work are summarised followed by a discussion of possible extensions.

### 6.1 Final remarks and contributions

In this thesis we have developed a method for estimating price sensitivity in situations where dependence between price and willingness to pay can be expected. A bivariate model was proposed to represent the customer's buying process and seller's pricing process, as well as the dependence between the two. The inclusion of instrumental variables allows the model to be identified, and the resulting estimates provide an accurate measure of the responsiveness of willingness to pay to changes in demand.

Whilst instrumental variable estimation has been used for endogeneity correction in pricing, this is typically performed in an ad hoc manner and without a formal model of the pricing process. By modelling price and willingness to pay jointly, the

relationship between them is decomposed into three components: (1) joint dependence on observed covariates, (2) dependence from unobserved covariates or endogeneity, and finally (3) the direct effect of price on willingness to pay. The extent to which dependence arises from unobserved covariates is measured by a correlation parameter, and positive price elasticity will be observed if this parameter exceeds the ratio of quote price and referral price standard deviation. As a diagnostic for positive price elasticity, we can check for positive correlation between the quote price residuals and willingness to pay.

Endogeneity in customised pricing can be viewed as an omitted variables problem; bias is caused by differences in expected willingness to pay across customers and sales agents' ability to identify these differences, based on information not available to the statistical modeller. A customer random effect for price is introduced to capture this information, and the customer effect is then entered into the probit regression as a predictor of willingness to pay. This goes some way towards alleviating any omitted variable bias.

Bayesian estimation of willingness to pay was performed, and formed the basis of a counterfactual analysis, in which the prices of previous quotes were modified to examine the impact on profit. This provides a means for the seller to improve their existing pricing policy, and it was found that agents are generally offering prices below their profit-maximising levels. Our analysis was limited to price changes of  $\pm 1\%$ , in order to avoid predicting outside the range of validity of the model, however it is also possible to maximise prices with the same approach.

## 6.2 Directions for future work

A limitation of the work carried out here is that estimation was performed in two stages, and willingness to pay is estimated conditionally on the estimated model for price. Holistic estimation of the model would incorporate information from both prices and customer responses. This is particularly pertinent to the customer random effect for price, which captures the agent's estimate of a customer's willingness to pay. Since the model includes non-nested random effects, estimation of the full model would require careful design of an MCMC scheme.

The extent of endogeneity bias provides a measure of the effectiveness of agents' pricing process, since a higher residual correlation between price and willingness to pay indicates that agents are able to more accurately identify the price sensitivity of customers. Using a hierarchical model, we could specify a unique residual correlation parameter for the quotes provided by each sales agent  $\rho_j (j = 1, \dots, n_a)$ , and in doing so, quantify how well each agent performs price discrimination. On this basis, the seller can determine which sales agents should be granted pricing discretion.

Zhang et al. (2014) note that in a market with volatile costs, reference price effects, or a customer's expectation of price play an important role in their decision to purchase. We have assumed here that a customer is aware of the spot price, or 'going price', at the time of the quote. A more realistic proposal may be to assume that customers' price expectations are formed based on previous price offers. Then for sales data in which customers make frequent purchases, referral price can be defined with respect to the change in price from the customer's previous price offer.

# Appendix A

## Parameter estimates

Table A.0.1: Quote price parameter estimates

Predictor		Subset 1 $\hat{\beta}_q$	Subset 2 $\hat{\beta}_q$	Subset 3 $\hat{\beta}_q$
Intercept		2.87 *** (0.054)	2.716 *** (0.056)	2.201 *** (0.093)
Volume	I(1000 ≤ vol < 1500)	-0.352 *** (0.004)	-0.242 *** (0.004)	-0.173 *** (0.009)
	I(1500 ≤ vol < 2000)	-0.41 *** (0.007)	-0.279 *** (0.007)	-0.406 *** (0.016)
	I(2000 ≤ vol < 2500)	-0.484 *** (0.006)	-0.306 *** (0.007)	-0.484 *** (0.015)
	I(2500 ≤ vol < 3000)	-0.533 *** (0.014)	-0.327 *** (0.014)	-0.531 *** (0.034)
	I(3000 ≤ vol < 4000)	-0.582 *** (0.016)	-0.35 *** (0.016)	-0.523 *** (0.035)
	I(4000 ≤ vol)	-0.558 *** (0.021)	-0.375 *** (0.02)	-0.628 *** (0.053)
Purchase History	0EEE	-0.086 *** (0.007)	-0.097 *** (0.008)	0.019 *** (0.016)
	1EEE	0.144 *** (0.005)	0.169 *** (0.005)	0.082 * (0.013)
	00EE	-0.163 *** (0.01)	-0.175 *** (0.01)	-0.039 (0.021)
	01EE	-0.008 *** (0.011)	0.02 *** (0.011)	0.054 * (0.022)
	10EE	0.056 ** (0.012)	0.069 (0.013)	0.026 (0.026)
	11EE	0.237 *** (0.006)	0.277 *** (0.006)	0.131 *** (0.014)
	000E	-0.235 *** (0.014)	-0.214 *** (0.015)	-0.07 (0.028)
	001E	-0.134 *** (0.019)	-0.098 *** (0.019)	-0.016 (0.034)
	010E	-0.034 *** (0.023)	-0.031 *** (0.023)	-0.058 (0.039)
	100E	-0.05 *** (0.025)	-0.087 *** (0.024)	0.018 (0.043)
	011E	0.087 *** (0.015)	0.132 (0.016)	0.031 (0.03)
	101E	0.095 ** (0.019)	0.119 * (0.019)	0.088 (0.039)
	110E	0.157 (0.017)	0.16 (0.018)	0.086 (0.038)
	111E	0.302 *** (0.007)	0.364 *** (0.007)	0.209 *** (0.016)
	0000	-0.301 *** (0.015)	-0.257 *** (0.015)	-0.098 (0.026)
	0001	-0.215 *** (0.02)	-0.168 *** (0.02)	-0.055 (0.033)
	0010	-0.146 *** (0.024)	-0.099 *** (0.024)	-0.131 *** (0.036)
	0100	-0.125 ***	-0.146 ***	-0.179 ***

Significance of associated p-values:

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , .  $p < 0.1$

Table A.0.1 – continued from previous page

Predictor		Subset 1 $\hat{\beta}_q$	Subset 2 $\hat{\beta}_q$	Subset 3 $\hat{\beta}_q$
	1000	(0.027) -0.139 *** (0.026)	(0.025) -0.216 *** (0.023)	(0.037) -0.032 . (0.042)
	0011	-0.053 *** (0.021)	0.053 *** (0.021)	-0.011 (0.032)
	0101	0.015 *** (0.027)	0.018 *** (0.027)	-0.042 * (0.04)
	1001	-0.03 *** (0.027)	0.066 *** (0.026)	-0.006 (0.046)
	0110	0.022 *** (0.027)	0.076 *** (0.027)	-0.05 (0.046)
	1010	-0.009 *** (0.029)	-0.041 *** (0.027)	-0.066 *** (0.047)
	1100	0.044 *** (0.027)	-0.006 *** (0.026)	-0.091 *** (0.053)
	0111	0.189 ** (0.014)	0.318 *** (0.014)	0.095 *** (0.023)
	1011	0.202 ** (0.016)	0.32 *** (0.016)	0.1 (0.031)
	1101	0.203 * (0.017)	0.304 *** (0.016)	0.026 * (0.036)
	1110	0.229 (0.016)	0.299 *** (0.016)	0.087 (0.035)
	1111	0.388 *** (0.007)	0.52 *** (0.008)	0.211 *** (0.016)
Quoted		1.103 *** (0.004)	0.597 *** (0.004)	0.49 *** (0.012)
Total Volume	log(volCount) ≤ 9.2	0.006 *** (0.004)	-0.035 *** (0.004)	-0.031 * (0.009)
	log(volCount) > 9.2	-0.072 * (0.006)	-0.057 *** (0.006)	0.075 ** (0.016)
Depot	1	-0.099 (0.041)		
	6	0.003 ** (0.035)		
	7	-0.202 *** (0.033)		
	9	-0.027 (0.038)		
	220	-0.015 *** (0.015)		
	230	-0.01 *** (0.016)		
	300	0.018 *** (0.023)		
	320	0.025 *** (0.031)		
	350	-0.199 *** (0.039)		
	470	-0.158 *** (0.035)		
	480	0.253 *** (0.034)		
	520	0.046 *** (0.022)		
	550	-0.616 *** (0.082)		
	620	-0.101 *** (0.037)		
	640	-0.176 *** (0.036)		
	670	0.014 *** (0.035)		
	690	0.25 *** (0.026)		
	720	-0.24 *** (0.021)		
	740	-0.255 *** (0.052)		
	820	0.116 *** (0.06)		
	840	-0.31 *** (0.037)		
	102		-0.354 *** (0.069)	
	170		-0.311 *** (0.044)	
	200		-0.234 *** (0.048)	
	280		-0.166 *** (0.026)	
	430		-0.097 *** (0.055)	
	450		-0.138 *** (0.031)	
	500		-0.208 *** (0.061)	
	580		-0.209 ***	

Significance of associated p-values:

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , .  $p < 0.1$

Table A.0.1 – continued from previous page

Predictor		Subset 1 $\hat{\beta}_q$	Subset 2 $\hat{\beta}_q$	Subset 3 $\hat{\beta}_q$
	590		(0.048) -0.248 ***	
	750		(0.037) -0.157 ***	
	770		(0.049) -0.175 ***	
	780		(0.05) -0.075 ***	
	790		(0.051) -0.287 ***	
	2		(0.049)	0.167 **
	240			(0.059) 0.326 ***
				(0.044)
Cost		0.252 *** (0.004)	0.181 *** (0.004)	0.111 *** (0.009)
Month	Feb	-0.067 *** (0.043)	-0.053 *** (0.036)	
	Mar	-0.077 (0.042)	-0.081 (0.035)	
	Apr	-0.306 *** (0.044)	-0.223 *** (0.037)	
	May	-0.424 *** (0.043)	-0.361 *** (0.036)	
	Jun	-0.539 *** (0.043)	-0.424 *** (0.037)	
	Jul	-0.639 *** (0.043)	-0.473 *** (0.037)	
	Aug	-0.581 *** (0.043)	-0.423 *** (0.036)	
	Sep	-0.506 *** (0.043)	-0.391 *** (0.036)	
	Oct	-0.398 *** (0.045)	-0.26 *** (0.038)	
	Nov	-0.249 *** (0.052)	-0.163 *** (0.043)	
	Dec	-0.128 *** (0.054)	-0.042 *** (0.046)	

Significance of associated p-values:  
 \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , .  $p < 0.1$

Table A.0.2: Referral price parameter estimates

Predictor		Subset 1		Subset 2		Subset 3	
		$\hat{\gamma}$	$\hat{\beta}_r$	$\hat{\gamma}$	$\hat{\beta}_r$	$\hat{\gamma}$	$\hat{\beta}_r$
Intercept		-2.985 *** (0.889)	-0.383 *** (0.116)	-3.173 *** (0.914)	-0.722 *** (0.068)	0.308 (0.305)	0.405 . (0.264)
Volume	I(1000 ≤ vol < 1500)	-0.258 *** (0.08)	-0.033 *** (0.008)	-0.329 *** (0.096)	-0.075 *** (0.005)	-0.641 *** (0.198)	-0.844 *** (0.252)
	I(1500 ≤ vol < 2000)	-0.24 *** (0.076)	-0.031 *** (0.009)	-0.262 *** (0.078)	-0.06 *** (0.005)	-0.691 *** (0.211)	-0.91 *** (0.267)
	I(2000 ≤ vol < 2500)	-0.416 *** (0.127)	-0.053 *** (0.013)	-0.4 *** (0.116)	-0.091 *** (0.006)	-0.92 *** (0.28)	-1.21 *** (0.355)
	I(2500 ≤ vol < 3000)	-0.46 *** (0.142)	-0.059 *** (0.018)	-0.506 *** (0.15)	-0.115 *** (0.011)	-1.021 *** (0.321)	-1.344 *** (0.402)
	I(3000 ≤ vol < 4000)	-0.57 *** (0.174)	-0.073 *** (0.021)	-0.563 *** (0.164)	-0.128 *** (0.012)	-0.982 *** (0.305)	-1.292 *** (0.381)
	I(4000 ≤ vol)	-0.574 *** (0.176)	-0.074 ** (0.025)	-0.634 *** (0.189)	-0.144 *** (0.015)	-1.265 *** (0.403)	-1.664 *** (0.499)
Purchase History	0EEE	-0.416 ** (0.135)	-0.053 * (0.029)	-0.495 *** (0.154)	-0.113 *** (0.018)	0.175 (0.138)	0.23 * (0.129)
	1EEE	0.918 *** (0.293)	0.118 ** (0.038)	0.747 ** (0.243)	0.17 *** (0.02)	1.366 ** (0.449)	1.797 *** (0.561)
	00EE	-0.749 *** (0.227)	-0.096 ** (0.035)	-0.718 *** (0.215)	-0.163 *** (0.02)	-0.144 (0.132)	-0.19 * (0.115)
	01EE	-0.102 . (0.073)	-0.013 (0.026)	-0.138 . (0.085)	-0.031 * (0.017)	0.447 * (0.194)	0.588 ** (0.215)
	10EE	0.363 ** (0.14)	0.047 . (0.029)	0.18 * (0.107)	0.041 * (0.018)	0.72 ** (0.269)	0.947 ** (0.319)
	11EE	1.16 *** (0.365)	0.149 *** (0.044)	1 *** (0.315)	0.228 *** (0.022)	1.677 *** (0.542)	2.206 *** (0.683)
	000E	-0.897 *** (0.27)	-0.115 ** (0.039)	-0.887 *** (0.263)	-0.202 *** (0.022)	-0.365 * (0.174)	-0.481 ** (0.177)
	001E	-0.496 *** (0.159)	-0.064 * (0.033)	-0.497 ** (0.161)	-0.113 *** (0.02)	0.005 (0.144)	0.007 (0.12)
	010E	-0.21 * (0.098)	-0.027 * (0.013)	-0.291 ** (0.121)	-0.066 *** (0.021)	0.158 (0.161)	0.208 . (0.147)
	100E	0.044 (0.09)	0.006 (0.013)	-0.095 (0.101)	-0.022 (0.021)	0.336 * (0.199)	0.442 * (0.2)
	011E	0.231 * (0.109)	0.03 (0.028)	0.151 . (0.103)	0.034 * (0.018)	0.824 ** (0.299)	1.083 ** (0.359)
	101E	0.615 ** (0.212)	0.079 * (0.036)	0.467 ** (0.178)	0.106 *** (0.021)	0.74 ** (0.282)	0.973 ** (0.329)
	110E	0.589 ** (0.205)	0.076 * (0.034)	0.529 ** (0.194)	0.12 *** (0.021)	0.97 ** (0.347)	1.276 ** (0.42)
	111E	1.291 *** (0.405)	0.166 *** (0.048)	1.119 *** (0.35)	0.255 *** (0.024)	1.69 *** (0.546)	2.223 *** (0.687)
	0000	-1.001 *** (0.298)	-0.129 *** (0.041)	-0.96 *** (0.279)	-0.219 *** (0.022)	-0.395 ** (0.17)	-0.52 ** (0.179)
	0001	-0.491 *** (0.157)	-0.063 * (0.033)	-0.527 *** (0.17)	-0.12 *** (0.021)	-0.015 (0.141)	-0.019 (0.117)
	0010	-0.546 *** (0.174)	-0.07 * (0.036)	-0.501 ** (0.167)	-0.114 *** (0.022)	0.062 (0.145)	0.082 (0.124)
	0100	-0.37 ** (0.134)	-0.048 . (0.035)	-0.355 ** (0.135)	-0.081 *** (0.022)	-0.139 (0.152)	-0.182 . (0.13)
	1000	-0.265 ** (0.112)	-0.034 ** (0.014)	-0.218 * (0.112)	-0.05 ** (0.021)	-0.18 (0.174)	-0.236 . (0.151)
	0011	-0.267 ** (0.104)	-0.034 (0.031)	-0.157 . (0.099)	-0.036 * (0.02)	0.376 * (0.189)	0.495 ** (0.203)
	0101	0.088 (0.097)	0.011 (0.013)	0.001 (0.105)	0 (0.022)	0.435 * (0.208)	0.572 ** (0.224)

Significance of associated p-values:

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , .  $p < 0.1$

Table A.0.2 – continued from previous page

Predictor		Subset 1		Subset 2		Subset 3	
		$\hat{\gamma}$	$\hat{\beta}_r$	$\hat{\gamma}$	$\hat{\beta}_r$	$\hat{\gamma}$	$\hat{\beta}_r$
	1001	0.291 *	0.037 **	0.186 .	0.042 *	0.521 *	0.685 **
		(0.137)	(0.014)	(0.127)	(0.022)	(0.243)	(0.265)
	0110	0.175 .	0.022 *	0.023	0.005	0.569 **	0.749 **
		(0.109)	(0.013)	(0.107)	(0.022)	(0.244)	(0.271)
	1010	0.214 *	0.027 *	0.133	0.03 .	0.293 .	0.386 *
		(0.121)	(0.014)	(0.119)	(0.023)	(0.194)	(0.189)
	1100	0.36 **	0.046	0.295 *	0.067 **	0.572 *	0.752 **
		(0.152)	(0.036)	(0.147)	(0.023)	(0.259)	(0.284)
	0111	0.634 **	0.081 **	0.378 **	0.086 ***	1.127 **	1.482 ***
		(0.216)	(0.033)	(0.153)	(0.018)	(0.382)	(0.473)
	1011	0.849 **	0.109 **	0.714 **	0.162 ***	1.191 **	1.567 ***
		(0.282)	(0.04)	(0.243)	(0.022)	(0.406)	(0.501)
	1101	0.974 **	0.125 **	0.729 **	0.166 ***	1.228 **	1.616 ***
		(0.316)	(0.043)	(0.249)	(0.022)	(0.42)	(0.517)
	1110	0.897 **	0.115 **	0.728 **	0.166 ***	1.247 **	1.641 ***
		(0.295)	(0.041)	(0.247)	(0.022)	(0.425)	(0.524)
	1111	1.471 ***	0.189 ***	1.25 ***	0.285 ***	1.916 ***	2.52 ***
		(0.461)	(0.053)	(0.388)	(0.025)	(0.615)	(0.777)
Quoted		0.975 ***	0.125 ***	0.854 ***	0.194 ***	1.399 ***	1.84 ***
		(0.29)	(0.028)	(0.245)	(0.012)	(0.426)	(0.543)
Total Volume	$\log(\text{volCount}) \leq 9.2$	0.314 ***	0.04 ***	0.369 ***	0.084 ***	0.261 **	0.343 ***
	$\log(\text{volCount}) > 9.2$	(0.011)	(0.108)	(0.007)	(0.087)	(0.107)	
		-0.129 ***	-0.017 **	-0.144 **	-0.033 ***	0.165 **	0.218 **
		(0.025)	(0.007)	(0.047)	(0.005)	(0.062)	(0.071)
Depot	1	0.277 ***	0.036 *				
		(0.087)	(0.016)				
	6	0.471 ***	0.06 ***				
		(0.141)	(0.018)				
	7	0.43 ***	0.055 ***				
		(0.13)	(0.016)				
	9	0.48 ***	0.062 ***				
		(0.147)	(0.019)				
	220	0.186 ***	0.024 **				
		(0.06)	(0.009)				
	230	0.346 ***	0.044 ***				
		(0.106)	(0.014)				
	300	0.284 ***	0.036 **				
		(0.086)	(0.012)				
	320	0.14 **	0.018 *				
		(0.046)	(0.008)				
	350	0.533 ***	0.068 ***				
		(0.16)	(0.02)				
	470	0.315 ***	0.04 **				
		(0.096)	(0.014)				
	480	1.943 ***	0.249 **				
		(0.618)	(0.084)				
	520	0.522 ***	0.067 ***				
		(0.158)	(0.019)				
	550	0.825 ***	0.106 **				
		(0.258)	(0.045)				
	620	0.428 ***	0.055 ***				
		(0.128)	(0.017)				
	640	0.149 ***	0.019 **				
		(0.047)	(0.008)				
	670	0.294 ***	0.038 **				
		(0.089)	(0.013)				
	690	0.25 ***	0.032 ***				
		(0.077)	(0.009)				
	720	0.093 **	0.012 .				

Significance of associated p-values:

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , .  $p < 0.1$



Table A.0.2 – continued from previous page									
Predictor		Subset 1		Subset 2		Subset 3			
		$\hat{\gamma}$	$\hat{\beta}_r$	$\hat{\gamma}$	$\hat{\beta}_r$	$\hat{\gamma}$	$\hat{\beta}_r$		
	740	(0.035) 0.363 ***	(0.009) 0.047 **						
	820	(0.11) 2.408 ***	(0.016) 0.309 ***						
	840	(0.74) 0.301 ***	(0.092) 0.039 **						
	102	(0.091)	(0.013)						
	170			0.784 *** (0.226)	0.178 *** (0.016)				
	200			2.147 *** (0.686)	0.489 *** (0.055)				
	280			0.711 *** (0.206)	0.162 *** (0.013)				
	430			0.115 *** (0.035)	0.026 *** (0.005)				
	450			0.703 *** (0.205)	0.16 *** (0.013)				
	500			0.199 *** (0.056)	0.045 *** (0.006)				
	580			0.172 *** (0.053)	0.039 *** (0.007)				
	590			0.079 ** (0.028)	0.018 *** (0.005)				
	750			0.235 *** (0.066)	0.053 *** (0.008)				
	770			0.205 *** (0.06)	0.047 *** (0.007)				
	780			0.396 *** (0.114)	0.09 *** (0.01)				
	790			0.517 *** (0.153)	0.118 *** (0.011)				
	2			-0.02 (0.024)	-0.005 (0.004)			1.242 *** (0.388)	1.633 *** (0.491)
	240							1.087 *** (0.328)	1.43 *** (0.418)
Distance	I(dist < 10)	0.424 *** (0.129)	0.054 *** (0.015)	0.365 *** (0.109)	0.083 *** (0.007)	0.064 * (0.038)	0.084 * (0.037)		
	$2.3 \leq \log(\text{dist}) < 3.7$	-0.125 *** (0.038)	-0.016 *** (0.004)	-0.092 *** (0.027)	-0.021 *** (0.002)	-0.025 ** (0.01)	-0.033 ** (0.011)		
Quote Int.	log(quoteInt)	-0.094 *** (0.027)	-0.012 * (0.007)	-0.039 * (0.021)	-0.009 * (0.004)	-0.16 ** (0.06)	-0.21 ** (0.071)		
	$4 \leq \log(\text{quoteInt}) < 5.6$	-0.084 *** (0.024)	-0.011 ** (0.004)	-0.091 ** (0.031)	-0.021 *** (0.006)	-0.109 * (0.051)	-0.144 ** (0.047)		
	$5.6 \leq \log(\text{quoteInt})$	-0.568 *** (0.179)	-0.073 ** (0.026)	-0.663 *** (0.203)	-0.151 *** (0.015)	-0.766 ** (0.259)	-1.008 *** (0.313)		
Customer	custEff	-0.37 *** (0.047)	-0.048 ** (0.02)	-0.416 *** (0.05)	-0.095 *** (0.013)	-0.719 ** (0.285)	-0.946 ** (0.356)		
Price	$0 \leq \text{custEff} < 0.3$	1.399 *** (0.25)	0.18 *** (0.052)	1.923 *** (0.356)	0.438 *** (0.036)	1.157 *** (0.13)	1.522 *** (0.118)		
Effect	$0.3 \leq \text{custEff}$	-0.687 *** (0.163)	-0.088 * (0.041)	-0.992 *** (0.216)	-0.226 *** (0.028)				
Significance of associated p-values: *** $p < 0.001$ , ** $p < 0.01$ , * $p < 0.05$ , . $p < 0.1$									

# Bibliography

- Agrawal, V. and Ferguson, M. (2007). Bid-response models for customised pricing. *Journal of Revenue and Pricing Management*, 6(3):212–228.
- Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature*, 19(4):1483–1536.
- Atchadé, Y., Fort, G., Moulines, E., and Priouret, P. (2009). Adaptive markov chain monte carlo: Theory and methods. Technical report.
- Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63(4):841–890.
- Bijmolt, T. H. A., van Heerde, H. J., and Pieters, R. G. M. (2005). New Empirical Generalizations on the Determinants of Price Elasticity. *Journal of Marketing Research*, 42(2):141–156.
- Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.

- Bowden, R. and Turkington, D. (1990). *Instrumental Variables*. Econometric Society Monographs. Cambridge University Press.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252.
- Boyd, D., Gordon, M., Anderson, J., Tai, C., Feng, Y., Kolamala, A., Cook, G., Guardino, T., Purang, M Krishnamurthy, P., Cooke, M., Nandiwada, R., Monteiro, B., and Haas, S. (2005). Manugistics Target Pricing System.
- Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, Cambridge, NY.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Cambridge University Press, 2nd edition.
- Collett, D. (2002). *Modelling Binary Data*. Chapman and Hall.
- Cramer, J. S. (2003). *Logit Models from Economics and Other Fields*. Cambridge University Press.
- Delattre, M., Lavielle, M., and Poursat, M.-A. (2014). A note on bic in mixed-effects models. *Electronic Journal of Statistics*, 8(1):456–475.
- Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression. i. *Biometrika*, 37(3-4):409–428.
- Elmaghraby, W., Jank, W., Karaesmen, I. Z., and Zhang, S. (2012). An exploratory

- analysis of B2B price changes. *Journal of Revenue and Pricing Management*, 11(6):607–624.
- Everitt, B. S., Landau, S., and Leese, M. (2011). *Cluster Analysis*. Wiley Publishing, 5th edition.
- Faraway, J. J. (2004). *Linear Models with R*. Chapman & Hall/CRC.
- Fong, N., Simester, D., and Anderson, E. (2010). Private label vs. national brand price sensitivity: Evaluating non-experimental identification strategies. Technical report, MIT.
- Garrow, L., Ferguson, M., Keskinocak, P., and Robert L, P. (2006). Expert opinions: Current pricing and revenue management practice across U.S. industries. *Journal of Revenue and Pricing Management*, 5(3):237–247.
- Gelman, A. (1996). Inference and monitoring convergence. In *Markov Chain Monte Carlo in Practice*, chapter 8. Chapman & Hall/CRC.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequence. *Statistical Science*, 7(4):457–472.
- Gilks, W., Richardson, S., and Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in Practice (Chapman & Hall/CRC Interdisciplinary Statistics)*. Chapman and Hall/CRC.
- Greven, S. and Kneib, T. (2010). On the behaviour of marginal and conditional aic in linear mixed models. *Biometrika*, 97(5):773–789.

- Gut, A. (2009). *An Intermediate Course in Probability*. Springer Texts in Statistics. Springer New York.
- Hausman, J. A., Leonard, G. K., and Zona, J. D. (1994). Competitive analysis with differentiated products. *Annales d'Économie et de Statistique*, (34):159–180.
- Hormby, S., Morrison, J., Dave, P., Meyers, M., and Tenca, T. (2010). Marriott International Increases Revenue by Implementing a Group Pricing Optimizer. *Interfaces*, 40(1):47–57.
- Hosmer, D. W. and Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in statistics-Theory and Methods*, 9(10):1043–1069.
- Hughes, A. W. and King, M. L. (2003). Model selection using aic in the presence of one-sided information. *Journal of Statistical Planning and Inference*, 115(2):397–411.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.
- Lehmann, E. and Casella, G. (2011). *Theory of Point Estimation*. Springer Texts in Statistics. Springer New York.
- Levene, H. (1960). Robust tests for equality of variances. In Olkin, I., editor, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pages 278–292. Stanford University Press.

- Müller, S., Scealy, J. L., and Welsh, A. H. (2013). Model selection in linear mixed models. *Statistical Science*, 28(2):135–167.
- OFT (2011). Off-grid Energy: An OFT market study. Technical Report October, Office of Fair Trading.
- Park, S. and Gupta, S. (2009). Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data. *Journal of Marketing Research*, 46(4):531–542.
- Petrin, A. and Train, K. (2003). Omitted Product Attributes in Discrete Choice Models. NBER Working Papers 9452, National Bureau of Economic Research, Inc.
- Petrin, A. and Train, K. (2010). A probabilistic choice model for market segmentation. *Journal of Marketing Research*, 47(February):370–379.
- Phillips, R. (2005). *Pricing and Revenue Optimization*. Stanford University Press.
- Phillips, R. (2010). Customized Pricing. Technical report, Columbia University Center for Pricing and Revenue Management.
- Phillips, R., van Ryzin, G., and Simsek, A. S. (2012). Endogeneity and Price Sensitivity in Customized Pricing. Technical report, Columbia University Center for Pricing and Revenue Management.
- Rivers, D. and Vuong, Q. H. (1988). Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models. *Journal of Econometrics*, 39:347–366.

- Robert, C. and Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer New York.
- Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling for various metropolis-hastings algorithms. *Statistical Science*, 16(4):351–367.
- Rossi, P., Allenby, G., and McCulloch, R. (2005). *Bayesian Statistics and Marketing*. Wiley.
- Sargan, J. D. (1976). Testing for misspecification after estimating using instrumental variables. Technical report, London School of Economics.
- Sherlock, C., Fearnhead, P., and Roberts, G. O. (2010). The Random Walk Metropolis : Linking Theory and Practice Through a Case Study. *Statistical Science*, 25(2):172–190.
- Sherlock, C., Xifara, T., Telfer, S., and Begon, M. (2013). A coupled hidden markov model for disease interactions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(4):609–627.
- Skrondal, A. and Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society*, 172(3):659–687.
- Stein, T. (2013). The Heat For B2B Marketing Is Rising. [http://www.cmo.com/articles/2013/5/2/the\\_heat\\_for\\_b2b\\_mar](http://www.cmo.com/articles/2013/5/2/the_heat_for_b2b_mar). Accessed: 2015-11-13.
- Talluri, K. and van Ryzin, G. (2004). *The Theory and Practice of Revenue Management*. Kluwer Academic Publishers.

- Taylor, A. E. (1952). L'hospital's rule. *The American Mathematical Monthly*, 59(1):20–24.
- Theil, H. (1961). *Economic forecasts and policy*. Contributions to economic analysis. Amsterdam, North-Holland, 2nd edition.
- Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, 2nd edition.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer.
- Villas-Boas, J. M. and Winer, R. S. (1999). Endogeneity in Brand Choice Models. *Management Science*, 45(10):1324–1338.
- Weiss, N., Holmes, P., and Hardy, M. (2005). *A Course in Probability*. Pearson Addison Wesley.
- Wood, S. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA.
- Wright, P. G. (1928). *The Tariff on Animal and Vegetable Oils*. MacMillan, New York.
- Yang, S., Chen, Y., and Allenby, G. M. (2003). Bayesian Analysis of Simultaneous Demand and Supply. *Quantitative Marketing and Economics*, 1(3):251–275.



- Zhang, J. Z., Netzer, O., and Ansari, A. (2014). Dynamic Targeted Pricing in B2B Relationships. *Marketing Science*, 33(3):317–337.